

Science and the Islamists

Muslim countries stand to gain much from science but will fail to do so if fundamentalists repress openness. Chronic neglect by Arab leaders doesn't help either.

Walk down a street in some predominantly Muslim cities and the chances are that you'll see taxis and buses bearing in Arabic the words "Seek knowledge, even as far as China". Attributed to the prophet Muhammad, the words encapsulate two principles: the duty of Muslims to seek an understanding of God's creation, and to search for knowledge beyond Islamic cultures.

In an unfortunate contrast, across the Muslim world secular governments are giving way to more overtly religious 'Islamist' leaderships that suppress free enquiry and critical-minded scholarship. How very different from the great age of scientific study lasting from the eighth to the thirteenth centuries, when leaders encouraged science and when debate and disagreement were more highly valued. That history can inspire today's young Muslims towards scientific ambition. But there is a danger of such inspiration being thwarted by currents in contemporary Islamic thought and politics.

The tension between support for science and restrictions on expression is one thread that runs through this special issue about science in Islamic countries (see page 19). Another is the shameful lack of interest in science displayed by wealthy Arab states, in contrast to Muslim states in east Asia and Africa. These themes emerged in an event that stimulated this issue: a meeting earlier this year organized by *Nature* at the Rockefeller Foundation's Bellagio Centre in Italy, attended by eminent scientists and science stakeholders from several Muslim countries. What appears this week is an exploration of the issues raised. International politics is avoided as a specific topic, even though it is unavoidable as a context. The intention, rather, is to survey some of the key influences and trends within Muslim states' attitudes towards, and uses of, science.

Another stimulus of our coverage is the diversity of the 57 Muslim nations that make up the Organization of the Islamic Conference (see the map on pages 20–21). This is manifest in their various states of economic health and development, as well as the degree of influence and fundamentalism of Islam.

Mistrust and isolation

Diversity is also apparent in the application of Islam to issues surrounding science, and indeed science itself. There are contradictory views, for example about the acceptability of weapons of mass destruction, therapeutic cloning, and the compatibility of evolution by natural selection with the existence of God. Without explicit guidance from the Koran or *sharia* law, the regulations are, as elsewhere, the outcome of religious opinion, culture and accepted evidence.

Unfortunately, diversity is less obvious in the tendency to restrict freedom of expression. It is disturbing that Islamist leaderships impose such restrictions while simultaneously encouraging science. It is reasonable to distrust their motivations when the originator and salesman of Pakistan's nuclear technology, Abdul Qadeer Khan, highlights the right of Muslim nations to have their own nuclear capability

independently of internationally agreed safeguards. The omens seem even more adverse given the bellicose rhetoric of the president of one of his major historical clients, Iran.

To what extent is science and scientific cooperation in these countries a casualty of these agendas? Certainly they cannot be helped by the resulting international mistrust and isolation of Islamist governments. The pursuit of collaboration creates not only opportunities but also risks for those who do it — and the international institutions that are meant to minimize those risks are weak.

Researchers in scientifically developed countries are further discouraged from contact because of politics, which has undermined non-political collaborations on synchrotrons and water purification, for example; because of regulations by non-Islamic states; and because of their own suspicions that collaboration is asking for trouble.

Collaboration and cooperation

There is no need to despair, however, and any tendency to do so should be challenged. The modern scientific world flourishes on collaboration and cooperation, wherever talent is to be found. The eleventh-century caliphates displayed such a spirit of engagement, and today's sultans and emirs need to foster it, seeking inspiration from both ancient and modern examples in Muslim states. For Western scientists, working with researchers who approach science's challenges within different cultures can only be good for science and encourage broader mutual awareness.

Another more philosophical and cultural motivation is a belief in the value of objective thinking in the midst of a maelstrom. There is much anti-Western feeling in Muslim states, and vice versa. There has never been a greater need for the measured, evidence-based approach to problems that comes from scientific training. Its contribution may be small amid the current turbulence, but it is all the more worth pursuing.

What are the critical components necessary for any Muslim nation with serious ambitions in science? A minimal requirement is an education system that embraces science as well as a critical approach, and at least one first-rate university. Appointments and promotions need to be transparently based on merit. It needs to call on talented expatriate scientists and duly reward them. And there must be openness to ideas and critical evaluation from other countries. The goals of developed scientific nations and individual scientists within them should be to foster these components at every opportunity.

If Islamists are willing to embrace *ijtihad* — unfettered reasoning — and critical investigation of the natural world, they could help unlock the great human potential of the Muslim world. This is not a question of simply importing Western ideas. They can draw inspiration also from the diverse attitudes of fellow Muslim states, reclaiming a great Islamic past in which new knowledge was valued and scholars were free to pursue all lines of enquiry.



A global call to arms

A seminal report on climate change deserves the world's attention.

It is fitting that Nicholas Stern, a British civil servant and former chief economist at the World Bank, should have presented his report on the economics of climate change at the Royal Society in London. For it was in a speech to the Royal Society 18 years ago, that Margaret Thatcher, then Britain's prime minister, became the first world leader to call for serious attention to be paid to the possibility of global warming.

That famous speech, although addressing science, was at heart political; so, in many ways, is Stern's review. It articulates a consensus on the need for action that Britain, thanks in part to Thatcher's remarks, has had longer to develop than many other nations. Its political nature should not, however, be seen as a flaw. Climate change is a matter of political choices, and the discussions that surround it must be political. Stern's review is an important contribution to that discussion (see pages 6–7).

Stern's review, which was commissioned by Gordon Brown, currently Britain's chancellor of the exchequer and in all likelihood its next prime minister, is almost as striking in its assumptions as in its conclusions. It insists, contrary to some utopian activists, that adverse consequences of climate change are unavoidable. In its choice of targets and assumptions, it tacitly supports the idea that the most feasible goal is one of stabilizing the level of greenhouse gas at roughly

twice the pre-industrial carbon dioxide level. Following through on the logic of this, Stern's review has almost as much to say about the economics of adapting to climate change as about the economics of mitigating it.

On mitigation, Stern concludes that the investment needed to control climate change is small compared with the costs of allowing it to continue unabated. The estimates of those costs are stated with more confidence than they would be in the scientific literature, and the review should not be taken as the last word on this. But the robust message is that the costs of mitigation, although vast, are manageable if efficient policies to put a price on carbon emissions are spread around the globe. The European emissions trading scheme might provide a nucleus for such pricing, although there are other possibilities.

The review also stresses the value of subsidies to help clean technologies establish themselves in the marketplace, and the importance of doubling research spending on energy issues. On both matters, its content should be noted by the people of California, who next week vote on Proposition 87, an initiative to establish a fund of \$4 billion (£2.1 billion) to spend on such support and research. A pre-allocated fund is not the ideal way to approach these issues, but it is the one on offer. Establishing it should help the development of clean technologies in California, and allow the state's citizens to declare their desire to move away from fossil-fuel dependency. It would therefore be a good proposition to pass. ■

"The investment needed to control climate change is small compared with the costs of allowing it to continue unabated."

Enough biodefence

Who wants a bioweapons lab next door?

Five years ago, in the aftermath of the attack on the World Trade Center, anthrax was found in the mail in several locations in the United States. The discovery heightened fears that the country was vulnerable to bioterrorist attack. Subsequently, the federal government has, according to the Center for Arms Control and Non-Proliferation in Washington DC, spent a staggering \$36 billion on biosafety-related activities.

Criticism of this approach was muted back in 2001, but five years on it is surging in the communities designated to host biosafety labs. Despite protests, however, the government is accelerating its investment in the biodefence complex that is now taking shape.

Last May, residents of Boston's Roxbury district sued the Boston University Medical Center and the National Institutes of Health (NIH) to try and block the construction of a \$178-million biodefence laboratory there. The lab, which will be equipped to perform experiments at biosafety level 4 — the highest classification of confinement — is already under construction. Last month, a district-court judge declined to halt the building work, but reserved the right to stop the project later on, subject to the results of an environmental review that the NIH has already agreed to perform.

The court's ruling suggests that the NIH and Boston University failed to adequately assess the environmental consequences of building such a lab in the middle of a large conurbation. A similar omission appears to have been made by the Department of Energy, which wants to construct a biosafety level 3 lab at the Lawrence Livermore National Laboratory in California — a site virtually surrounded by suburban sprawl. An appeal court there ruled last month that the energy department has yet to conduct a sufficiently thorough assessment of the laboratory's risks, particularly regarding the risk of a terrorist attack on the site itself.

These cases follow protests against a new biosafety level 4 lab being built by the Department of Homeland Security (DHS) at Fort Detrick, Maryland, as part of a \$105-million National Biodefence Analysis and Countermeasures Center. And they offer a taste of what's to come with another DHS project, a planned \$450-million National Bio and Agro-Defence Facility, as yet unlocated.

These objections have concentrated on local questions, but the proliferation of such labs begs the broader question of how much biodefence is too much. In the febrile climate of 2001, the Bush administration wasn't pressed sufficiently to explain why such proliferation of knowledge is in the national interest. Five years on, the time has come for it to do so. ■

"The proliferation of such labs begs the broader question of how much biodefence is too much."

RESEARCH HIGHLIGHTS

Darwinian serenade

Proc. R. Soc. Lond. B doi:10.1098/rspb.2006.3736 (2006)

A population of Amazonian frogs that woos with calls (*Physalaemus petersi*, pictured) seems to be splitting into several species as the females' preferences for different types of call diverge, report researchers at the University of Texas in Austin.

W. Chris Funk, now of the United States Geological Survey in Corvallis, Oregon, and his colleagues gathered genetic evidence to support the idea. They found, for example, that gene flow between groups that have a simple call composed of a single whine component (which Funk pronounces "Chu! Chu!") and groups that have a complex call containing a whine and a squawk ("Chu hanh! Chu hanh!") is 30 times lower than within these groups.

Speciation by sexual selection, although supported by theory, is hard to catch in action.



W. C. FUNK

MICROFLUIDICS

Go with the flow

Lab Chip 6, 1277-1278 (2006)

The controlled flow of liquids in microscopic channels on 'microfluidic' chips could permit chemical analysis of tiny samples as well as more efficient industrial chemical synthesis. But how can the fluid traffic be directed down the right channels?

Microscopic valves and gates are cumbersome, so Nathaniel Robinson and his colleagues at Linköping University in Sweden propose to do away with all moving parts.

They made channels with floors whose wettability can be controlled electrically, using conducting polymers that have different surface properties when electrochemically oxidized or reduced. Water injected into such a system flows preferentially along the oxidized channels (pictured right).

CHEMISTRY

Scenting money

Angew. Chem. Int. Edn. Engl. 45, 7006-7009 (2006)
"Money doesn't stink", the Roman Emperor Vespasian's son allegedly told him. But your hands do smell after handling coins. Surprisingly, the reason for this has only just been determined.

Dietmar Glindemann of Virginia Polytechnic Institute and State University, Blacksburg, and his co-workers have found that the characteristic metallic aroma produced by touching iron and copper isn't

intrinsic to the metals at all — it is a type of body odour. These metals rapidly reduce lipid-based compounds at the skin surface to aldehydes and ketones that give the distinctive musty smell we associate with metal. This process might also explain the metallic taste often experienced in drinking water, and the metallic smell of blood.

ASTRONOMY

Missing helium

Science doi:10.1126/science.1133065 (2006)

Low-mass giant stars may have been spewing much less helium-3 into space than was previously thought, a new model suggests. This could explain why cosmic levels of helium-3 are lower than theory predicted.

Helium-3 was produced in the Big Bang and has since been emitted by giant stars, but the amount observed in the universe

at large is less than had been expected given theoretical models of these two processes. Now, Peter Eggleton and his colleagues at the Lawrence Livermore National Laboratory in California have seen the fate of the missing helium in a computer model of a red giant star.

The researchers' three-dimensional model revealed a subtle instability that leads to turbulence within the star and a greater rate of helium-3 burning, greatly reducing the amount lost to space. This makes cosmic levels not much higher than those predicted by Big-Bang theorists seem reasonable.

CELL BIOLOGY

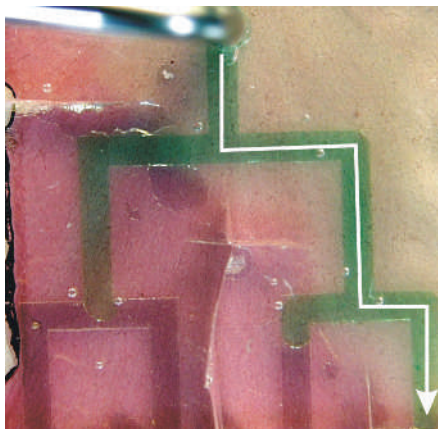
Housekeeper has two jobs

Neuron 52, 321-333 (2006)

A novel player in the business of helping brain cells to communicate with each other has been discovered by Deborah Nelson at the University of Chicago, Illinois, and her colleagues.

All nerve cells have channels in their membranes that allow ions to pass in and out of the cell. Ion transport through calcium and potassium channels, for example, triggers nerve cells to fire impulses. Chloride channels, on the other hand, were thought to have a 'housekeeping' function, preserving cell volume and maintaining electrical charge.

Nelson's group found that a chloride channel known as ClC-3 amplifies cell-to-cell communication in immature neurons by



L. ROBINSON

propagating action potentials. Later, as the neuron matures, its role changes to one that dampens interneuronal signalling.

MEDICINE

Targeting birth control

Nature Med. doi:10.1038/nm1420 (2006)

In the quest for a male contraceptive, the drug Adjudin initially seemed a promising candidate. But development slowed when researchers found that the dose that inhibited fertility also caused muscle atrophy and liver inflammation in three out of ten male rats.

Now, Chuen-yan Cheng of the Population Council in New York and his colleagues have targeted Adjudin directly to the testes. They did this by coupling it to a recombinant form of a hormone — known as follicle-stimulating hormone — that is important for producing sperm and whose receptor is restricted to the type of cell in the testis known as a Sertoli cell. Targeting the drug allowed them to block fertility in rats using doses of Adjudin a thousand-fold lower than previously recommended. The researchers hypothesize that targeting Adjudin could eliminate its unwanted side effects.

PHYSICS

Cool it

Nature Phys. doi: 10.1038/nphys443 (2006)

For the first time, physicists have cooled a gas by demagnetizing its atoms. This simple cooling technique has previously only been used for solids.

Marco Fattori and his colleagues at the University of Stuttgart in Germany used lasers to trap a magnetically polarized gas containing about a million chromium atoms. When the team ramped down the applied magnetic field, the atoms' magnetic alignments spread into a number of energy levels, absorbing thermal energy in the process. The temperature of the gas dropped from 19 microkelvin to 16.5 microkelvin in 5 seconds.

By using this process to ratchet the temperature lower still, the authors believe they can push the atoms into a quantum state known as a Bose–Einstein condensate.

BIOLOGICAL CHEMISTRY

New twist for DNA

J. Am. Chem. Soc. doi:10.1021/ja065606n (2006)

Unnatural DNA bases that fluoresce will pair up with DNA's natural bases to form a double helix, report Eric Kool and his colleagues at Stanford University in California. Strands of such bases could be used to image and sense specific DNA sequences.

The researchers deciphered the structure of a stretch of 'xDNA' ten bases long in which each of the four natural DNA bases was matched to one of four 'expanded' unnatural bases containing a benzene ring.

The modified DNA structure (pictured left) is similar to the original form, with the bases forming hydrogen-bonded pairs in a right-handed helix. But the molecule's backbones are further apart and the helix turn is slower, with more base pairs per turn, than in natural DNA.

CELL BIOLOGY

Dual use

Cell 127, 397–408 (2006)

The protein PGC-1 α not only stokes up energy production in mitochondria, but also induces enzymes to detoxify the damaging reactive oxygen species (ROS) that are a by-product of energy production.

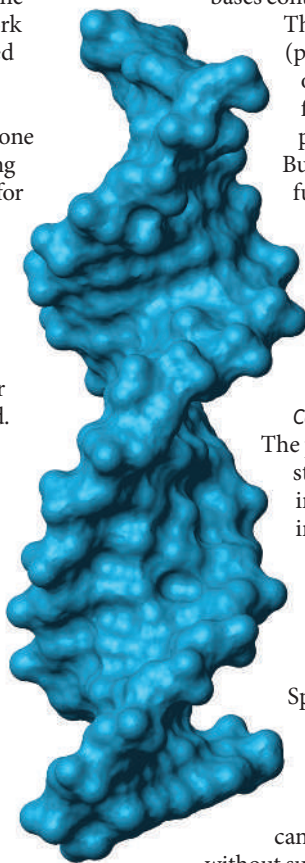
This finding, from Bruce Spiegelman of Harvard Medical School in Boston, Massachusetts, and his colleagues, might explain how some tissues, such as muscle, can ramp up mitochondrial activity

without suffering self-inflicted damage.

The researchers showed that mice lacking the *Pgc-1 α* gene are abnormally sensitive to ROS damage, and that PGC-1 α protects brain cells in culture from ROS damage. They propose that PGC-1 α could be a therapeutic target for diseases such as Alzheimer's and Parkinson's, which are associated with mitochondrial dysfunction and ROS damage.

Correction

The reference in 'The Big Draw' (*Nature*, 443, 484; 2006) should have been given as *Angew. Chem. Int. Edn Engl.* doi: 10.1002/anie.200603142 (2006).



S. R. LYNCH/STANFORD UNIV.

JOURNAL CLUB

Euan Nisbet
Royal Holloway, University of
London, UK

Methane humour hides serious issues, argues a geologist.

Methane people are like the Jumblies of Edward Lear's poem — far and few, far and few. We work together across continents, sharing air samples in 'round robin' experiments. We also suffer bovine eructation jokes; but it's not just about cows' breath.

Methane is a potent greenhouse gas, produced by wetlands, fossil fuels, grass and forest fires, and rice paddies. But there are major puzzles in the global budget. Methanologists at a recent meeting in Cape Town, South Africa, discussed two of these.

One stems from experiments suggesting that plants emit methane (F. Keppler *et al. Nature* 439, 187–191; 2006). Our first reaction was 'surely not!', yet satellite studies of tropical forests seem to back up the result. We folk who once energetically sampled the atmosphere on mountain peaks are becoming plant biochemists, watching air bubble through flasks.

Another dispute surrounds methane leaks from fossil fuels. We used to estimate leakage by measuring the carbon-14 in methane. But nuclear power stations release this, so the technique became useless.

Recently, researchers reanalysed methane's carbon-14 record in a way that avoids the nuclear problem (K. Lassey *et al. Atmos. Chem. Phys. Disc.* 6, 5039–5056; 2006). The result is surprising: roughly 29% of atmospheric methane is fossil — much more than expected. Some of this leakage is geological (that's a worrying puzzle too), but most is from gas and coal industries.

Methane is the quickest, cheapest, easiest greenhouse-gas target. But governments aren't interested. Important monitoring programmes were cut recently in Europe and Australia, and UK regulators even limit industry proposals for leak reduction. Are we looking a gift cow in the mouth?

SPECIAL REPORT

How much will it cost to save the world?

The *Stern Review* won't be the last word on the cost of global warming. But it has upped the stakes in the debate. **Jim Giles** reports.

He's a highly respected researcher and a former chief economist at the World Bank. He had a year and the help of more than 20 of Britain's brightest civil servants and academics. His work was commissioned by Gordon Brown, who controls Britain's budget and is likely to be the country's next prime minister. So could Nicholas Stern settle the debate about the economic impact of climate change?

No chance. The *Stern Review on the Economics of Climate Change*, released on 30 October, has been praised by many economists, who say it sets a new benchmark for quality and thoroughness. But its dramatic conclusions, including the claim that tackling climate change would cost 20 times less than doing nothing, were immediately attacked by right-wing commentators and other economists. Some add that the report covers such complex ground that it should be seen as a political rather than a scientific document (see Editorial, page 2). Just as arguments about the validity of climate science are dying down, it seems that a new battle is looming over how much the world should spend to tackle climate change.

At first glance, Stern's review provides ammunition for those who advocate tough measures. Stern predicts that if greenhouse-gas emissions carry on as normal, between 5% and 20% could be wiped off the global gross domestic product (GDP) by the beginning of the next century. "This allows us to put a dollar number on the cost of climate change for the first time," says Philip Clapp at the National Environment Trust, a not-for-profit group in Washington DC.

Equally important, adds Clapp, Stern helps to counter campaigns by the oil and coal industries, which have highlighted the costs of tackling climate change. The review says that stabilizing greenhouse-gas concentrations at roughly double pre-industrial levels would cost a relatively paltry 1% of GDP, so it provides strong support for measures such as mandatory emission limits and public investment in green technologies. The key message is that acting now will cost far less than acting later.

The report has prompted some alarmist headlines, with one British newspaper declaring "Act now or the world we know will be lost forever". Yet the methodology used is first-rate, say supporters. Estimating costs over decades, and trying to factor in social and economic changes that will take place, is enormously tricky. But the report contains a comprehensive meta-analysis of existing studies on mitigation costs. It also includes new results from a model that builds on the handful of previous studies that have tried to calculate the impact on global GDP.

Inevitable criticism

The upper bound of 20% for loss of GDP, which is higher than previous estimates, is due to several factors. Stern generally uses low numbers for discount rates, the parameter used by economists to compare future and current costs. His team also considered a range of values for climate sensitivity — the global temperature rise caused by a given increase in greenhouse gas concentrations. Using a range leads to the higher upper limit for GDP reduction, but the results are more realistic.

"This is unquestionably head and shoulders above previous economic assessments," says Michael Grubb, an energy economist at Imperial College London, who contributed to the review. But others have homed in on the many assumptions that had to be made. Many right-wing commentators attacked the review on these grounds, some even starting work before publication. The use of one global development scenario, in which population reaches what demographers say is an unrealistic figure of 15 billion, attracted criticism.

Economists were also quick off the mark. In a four-page critique compiled within hours of the review's publication, economist Richard Tol of Princeton University accused Stern of selective reporting: "For water, agriculture, health and insurance, the Stern Review consistently selects the most pessimistic study in the literature."

On sea-level rise, for example, Tol says the review underplays the role of better coastal defences. Roger Pielke Jr, an expert in climate-change policy at the University of Colorado,

"In a sense it's neither here nor there whether you believe the numbers."

Money man: Nicholas Stern's review puts a price on climate change.

Boulder, also accuses Stern of "cherry picking" alarming results from the literature on the link between natural disasters and climate change.

Such criticisms come as no surprise to Mike Hulme, director of the Tyndall Centre for Climate Change Research in Norwich, UK. Hulme says that the British government has asked him many times to conduct a study on the total cost of climate change. He declined, as he does not feel it's a question that researchers can answer. Difficulties in estimating the impact of strategies such as coastal defences are only part of the problem. When other assumptions, such as the economic cost of species extinctions, are included, Hulme feels that the uncertainties become so great that he would not be able to defend the end result.

He says that Stern's team seems to have done a good job, but with so many assumptions involved, and the review having been conducted by a political appointee: "This is not the last word of scientists and economists, it's the last word of civil servants."

Trillion-dollar questions

An academic involved in the Stern review, who did not want to be named as he was speaking on behalf of the team, told *Nature* that each assumption is based on the "sound principles of science and economics" and that the review spells out how the uncertainties affect the final



result. He adds that the involvement of civil servants does not justify the “dangerous and incorrect” allegation that the report is politicized: “There was never any political pressure to produce high numbers.”

The inevitable criticisms of the review need not reduce its impact. Negotiations over the future of the Kyoto Protocol are taking place this week (see ‘Kyoto looks to the future’), and the review will strengthen the case of countries like Britain, which want tough limits on future emissions. Stern discusses several ways in which this could be done; one suggestion, that the current European Union emission-trading scheme be made global, was backed by Brown at the review’s launch. Stern also notes that a substantial hike in public energy-research spending is needed, as market factors alone will not drive the development of technologies such as improved solar cells and biofuels that are needed to reduce emissions.

Stimulating debate over spending decisions, rather than putting a figure on the true costs of climate change, may be the review’s main legacy. Although Hulme questions the assumptions behind the headline result, he has no doubt that Stern is broadly correct: “In a sense it’s neither here nor there whether you believe the numbers. This will take the discourse away from the costs of taking action and put attention onto the costs of inaction.” ■

Kyoto looks to the future

Next week, thousands of delegates from more than 180 countries will fly to Nairobi, Kenya, for talks about the Kyoto Protocol on climate change. Is this an exercise in futile bureaucracy, as some critics claim, or will it break new diplomatic and economic ground?

Kyoto has polarized opinion since its conception. The treaty calls for signatories to make an average 5% reduction (relative to 1990 levels) of greenhouse-gas emissions by 2012. Only developed countries are required to make cuts, so major developing nations such as China, India and Brazil are not included. In 2001, the United States and Australia dropped out, claiming that the protocol would overburden their domestic economies.

The treaty finally came into force in 2005, when Russia signed up and the countries committing to targets accounted for 55% of 1990 emissions.

The first period during which countries’ emissions will be measured and assessed is 2008–12. But several mechanisms are already in place. Europe started a mandatory emissions-trading system in January 2005, allowing carbon-hungry companies to buy credits from companies with emissions to spare. It’s unclear whether the scheme has lowered emissions, and critics say the carbon allowances were set too high. But the principle is seen as a success, and the market could become the nucleus of a global emissions-trading scheme.

Kyoto also gives credits to industrialized countries for helping poorer countries reduce or lock away emissions, for example by funding alternative energy or reforestation projects. The Clean Development

Mechanism (CDM), launched last year, covers developing countries, and the Joint Implementation (JI) mechanism, which covers projects in other industrialized countries such as central and eastern Europe, launched last week.

These systems aim to force rich countries to implement real carbon reductions. But critics complain that they allow countries to avoid improving energy efficiency at home, and that the projects target easy savings rather than encouraging investment in new technologies.

Still, progress so far looks impressive. Around 4,000 CDM and JI projects are

“Canada, Japan and several EU members are lagging behind, and face an enormous effort to catch up.”

planned or in place, mostly funded by private investors in the European Union (EU) and Japan. Not all will work out, but if they did they could save the equivalent of 2.2 gigatonnes of carbon dioxide — around seven times the amount that the EU needs to save by 2012.

These mechanisms will be needed. Many Kyoto countries, including Canada, Japan and several EU members, are lagging well behind target, and face an enormous effort to catch up.

The reasons vary. Japan’s growing industry is emitting 8% more greenhouse gases than in 1990, so now needs a 14% cut by 2012. In Canada, changes of government and the desire to profit from oil deposits in Alberta, have led to Kyoto slipping off the agenda. In March, it became the first country to say that it probably will not meet its 2012 target.

Italy, Spain, Austria and Finland are also having trouble, so as the 2012 date gets closer, many will be looking to buy credits to make up the shortfall. One potential bump in the road is that Russia and Ukraine, whose emissions went down sharply after the political and economic collapse of the Soviet Union, are likely at some point to sell their surplus allowances. If they sell too much too fast, the price of carbon could drop drastically.

Another problem is that it is not clear how Kyoto will determine who has met their targets, and what happens to countries that have not. The rules say that countries missing their targets will be suspended from carbon trading and must make up the shortfall, plus a 30% penalty, in the second commitment period after 2012. But the Kyoto compliance committee has not yet decided how to measure countries’ carbon emissions, or how to enforce measures on those that do not comply. It has not even issued an early warning to countries running the risk of falling short.

With so much uncertainty, it is unclear whether countries will have an incentive to make the tough decisions needed to hit targets. Emission assessment and compliance will both be discussed in Nairobi next week. But a post-2012 negotiating session is not even on the agenda.

Fed up with the murkiness, some climate-policy analysts now argue that the Kyoto Protocol is not worth continuing with, preferring to look towards a new type of agreement that would have the support of more regions and countries, especially the United States. But whether it survives or not, the treaty is set to provide a solid foundation for future agreements on climate. **Quirin Schiermeier**



MEDICS ON TRIAL
Follow events in Libya, where medics stand accused of infecting children with HIV.
www.nature.com/news

M. TURKIA/AFP/GETTY

ON THE RECORD

“Queen Elizabeth has 10 times the lifespan of workers and lays up to 2,000 eggs a day.”

Did an over-enthusiastic spell-check cause a Reuters story on the honeybee genome, published in *Nature* last week, to go somewhat beyond the scientists' findings?

“Everything from making a fighter more fuel-efficient to looking at the materials that munitions are made of.”

Spokeswoman Deborah Allen explains how arms-maker British Aerospace is investing in more environmentally friendly weapons.

“Money was spent on the Russian mafia as we tried to clone mammoths. You can't say that so we expensed it as money for cows.”

Woo Suk Hwang finally takes the stand in his defence (see page 12).

ZOO NEWS**Hot hedgehogs**

As the British get ready for 5 November celebrations of bonfires and fireworks, the Hedgehog Preservation Society is pleading for people to check bonfires before lighting

them. “Piles of bonfire material look like five-star hotels to a hedgehog,” it warns.

**OVERHYPED**

It is dogma in the public-health community that the stigma of being HIV positive fuels the epidemic. The theory goes that people are afraid to use condoms, say, or seek testing, for fear that others might suspect their HIV status. But in a provocative essay in *PLoS Medicine*, Daniel Reidpath and Kit Yee Chan say there's no evidence for this. Establishing the link requires longitudinal data on levels of stigma and rates of infection, or at least a correlation between them in different areas.

Sources: BBC, Reuters

Ethiopian plan for Lucy tour splits museums

Lucy, the famed Ethiopian fossil, may soon take to the sky to sparkle on her first US tour. But where the 3.2-million-year-old hominid will travel after an initial appearance in Houston is unclear, with many anthropologists fearing that the risk to the fragile skeleton is too great.

On 24 October, Ethiopian government officials announced an agreement with the Houston Museum of Natural Science to exhibit Lucy's skeleton and nearly 200 other artefacts, beginning in September 2007. Few have seen the Lucy remains, found in 1974 and named after the song *Lucy in the Sky with Diamonds* by The Beatles. The skeleton, which is 40% complete, is kept in a vault in the basement of the Ethiopian National Museum in Addis Ababa.

If palaeoanthropologists agree on anything, it is that original fossils should not travel, to keep them safe and available for research. A policy statement issued by the International Association for the Study of Human Paleontology says that valuable remains should be moved only for “compelling scientific reasons”. Casts are typically displayed instead.

But Joel Bartsch, president of the Houston museum, argues that museums such as his have the expertise to handle delicate specimens. He sees the exhibit as an opportunity “to raise the profile of a country that is home to the cradle of mankind”.

Sating public hunger for historical reality can bring sorely needed funds to the home countries of valuable artefacts. The Tutankhamun tour, currently being visited by throngs across the United States, has brought Egypt millions of dollars to help it preserve its antiquities. But whereas Egypt has kept the most valuable human remains at home, Ethiopia and Kenya seem willing to tour their prize exhibits.

In another attempt at a high-profile fossil tour, officials from the National Museums of Kenya flew to the Field Museum in Chicago in September to propose an exhibit centred on Nariokotome Boy, a 1.6-million-year-old fossil

of a near-complete *Homo erectus* youth.

The Kenyans then announced that the exhibit would lead to the return of some Kenyan antiquities held in the Field Museum — notably the Tsavo lions, stuffed man-eaters famous a century ago for killings outside Mombasa. But Neil Shubin, provost of the Field Museum, says he was stunned. “There was never a formal proposal made for an exhibit,” he says, adding that there is no plan to send the lions back.

Details of the Lucy tour, including where it will appear and how much money Ethiopia will get, are also unclear. Major US museums, including the American Museum of Natural History in New York and the Smithsonian Institution, say they are not interested, out of concern for the remains. Bartsch admits that no other museums have signed up yet, but says he anticipates intense interest for a ten-city tour.

Jara Haile Mariam, manager of the Ethiopian culture ministry's Authority for Research and Conservation of Cultural Heritage, says that Ethiopia hopes to get US\$5–6 million from the tour, with

museums paying about \$300,000 each for four-month exhibitions, plus a percentage of the takings. The Ann and Robert H. Lurie Foundation is also reported to have pledged \$1 million, but Bartsch and the foundation declined to comment on the financial details.

Despite the difficulties over agreeing terms, those who advocate sending African fossils abroad say it will allow them to be studied at high-tech US institutions, and will encourage tourism and trade in their countries of origin by inspiring Western museum-goers.

But many palaeoanthropologists are concerned by the trend. “This makes no sense,” says Meave Leakey, whose husband Richard's team found the Nariokotome Boy in 1984. She argues that there are already adequate resources for analysis in Africa. As for increasing tourism: “This is the reverse of what will logically happen. If the fossils can be viewed in the United States, why travel to Ethiopia or Kenya?”

Rex Dalton



Stay at home: researchers say that Lucy is too fragile to travel.

T. MCHUGH/SPL

S. SHOSTAK/SETI INST.



The Allen Telescope Array will consist of 350 satellite dishes — if funds can be found.

Search for alien signals stalls for want of cash

An ambitious radio array project that will join the search for extraterrestrial intelligence (SETI) is running into money problems.

Construction of the Allen Telescope Array, named after its chief benefactor, Microsoft co-founder and billionaire Paul Allen, will halt at the end of this year unless further funding is found. Allen is currently withholding millions because the project has failed to recruit other donors.

"Down here in the trenches it's really terribly worrisome," says Jill Tarter, director of the privately funded SETI Institute in Mountain View, California, which is developing the project.

The array, which is being built at Hat Creek Radio Observatory in California, was conceived in the late 1990s as a cheap way to search for extraterrestrial radio transmissions. For about \$25 million, researchers believed they could build a radio telescope with 350 commercially available satellite dishes. The price went up to \$43 million in 2003, when scientists decided to upgrade the dishes and other components, allowing the array to do radio astronomy as well as searching for alien signals.

In 2000, Allen gave the project's research budget \$11.5 million. He pledged \$13.5 million more for construction in 2003 — but the money was contingent on the SETI Institute raising another \$16 million in private funding (see *Nature* **428**, 358; 2004).

To date, the institute has raised less than \$9 million. Allen is withholding \$3.85 million until the institute can "meet its contractual obligations", says Jason Hunke, a spokesman

for the Paul G. Allen Family Foundation, based in Seattle, Washington.

Astronomers often use private funding for telescopes, and almost as often find themselves in tight spots. Researchers at the California Institute of Technology (Caltech) in Pasadena nearly abandoned the second of the huge Keck telescopes before NASA provided a multi-million-dollar bailout.

And the Sloan Digital Sky Survey, a \$70-million map of the sky, had to beg and borrow from collaborators, says James Gunn, a founder of the project at Princeton University in New Jersey. "We were on the edge of financial disaster more or less continuously," he says, adding that the project has since repaid most of its debts.

With the funds currently available, the Allen Telescope Array will have just 42 of its planned 350 dishes up by the end of the year, according to Leo Blitz, an astronomer at the University of California, Berkeley, the project's major collaborator. This will still be enough to start doing science, he adds.

But the full array would be much better, says Shri Kulkarni, a radio astronomer at Caltech. The more expensive array would deliver high-resolution, wide-angle images of the sky relatively cheaply, he says.

And it would be an important proof-of-principle for future radio observatories, such as the international Square Kilometre Array. If the Allen array didn't move forward, Kulkarni says, "it would be a pity in almost every way I could think of".

Geoff Brumfiel

"Down here in the trenches it's really terribly worrisome."



HOOPS, SWEAT AND TEARS

Physicists help a US basketball team get to grips with its new ball.

www.nature.com/news

O. FRANKEN/CORBIS



Rosy prospect? A compound found in red wine might help to protect against the health problems associated with obesity.

A votre santé: now in pill form?

David Sinclair believes resveratrol is a miracle drug. He's been taking it for three years because he hopes it will help him live a healthier life, despite a lack of evidence that it works in humans — or any data on the safety of long-term exposure.

But this week, Sinclair comes a step closer to proving that he's on to something. He and his colleagues report in a paper published online in *Nature*¹ that this compound counteracts the ill effects of a high-fat, high-calorie diet — at least in mice. To some scientists, the finding has the whiff of a landmark discovery: it could lay the foundations for a future drug that blocks the toxic effects of obesity in humans. But how solid is that foundation?

The research is certainly striking. The study was led by Sinclair, of Harvard Medical School in Cambridge, Massachusetts, and Rafael de Cabo of the National Institute on Aging in Baltimore, Maryland. Their team divided one-year-old mice into three groups. One group ate regular food. Another ate high-calorie food. The third group ate the high-calorie diet along with a daily dose of resveratrol, a chemical found in many plant species, including red grapes — and thus also red wine.

After six months, all the mice fed high-calorie diets had grown fat. But after a year the mice in the resveratrol group seemed a lot healthier than their high-calorie-fed counterparts. The drug seemed to prevent these mice from developing a diabetes-like illness and liver damage, and reduced their risk of death by 30%,

compared with the mice on high-calorie diets that did not get the drug.

That's a potentially revolutionary finding. But for those desperate to undo the effects of years of overeating, it is a far cry from a human cure. First, the study doesn't attempt to show whether resveratrol reverses the damage caused by the toxic diet, as the mice were on the drug when they started the diet. The equivalent experiment would be to feed a human a relatively healthy diet until middle age, and then force him or her to binge on Big Macs and resveratrol simultaneously.

Second, there is a dearth of data on how resveratrol works in mice — and no evidence that it works at all in humans. Sinclair believes that it triggers the same pathways as those activated by calorie-restriction diets, which have been shown to increase lifespan in some animals (but not in humans, yet). He hypothesizes that resveratrol works through proteins called sirtuins. But although there are data to prove this in yeast, worms and fruitflies, no analogous data have proved it for humans or mice — despite tantalizing hints. So it is hard to know how well resveratrol's benefits will apply in people.

A third caveat is that mice on resveratrol didn't exactly mimic people on calorie-restriction diets. The mice stayed fat, for instance, and their cholesterol levels were far higher than those of mice on the standard diet. This means either that Sinclair and de Cabo decoupled

obesity from its downstream ill-health effects or that resveratrol doesn't really mimic calorie restriction in mammals. Of course, the mechanism isn't so important if the drug works. But it means that, for now, experts are cautioning against the idea of rushing to an Internet pharmacy to buy resveratrol². It is a dicey idea to take lifelong doses of a drug without having a clue about its mechanism of action or its long-term effects on the body.

Finally, although the study itself is robust as published, the analysis involves only 121 mice, many of which are still alive — the experiment isn't truly over until all the mice have died. Further studies by the same group will look at whether resveratrol extends lifespan in the healthy mice, too. Until all these results are in, we don't really have a complete picture of the drug's effects in obese and normal mice.

Even so, Sinclair is forging ahead to answer some of these questions. He is a co-founder of Cambridge-based Sirtris Pharmaceuticals, which aims to develop drugs that act on sirtuins. This summer, Sirtris tested a modified version of resveratrol in 85 healthy men. The company said on 4 October that it saw no ill effects of the drug in these volunteers, so it has now begun a clinical trial in 90 patients with diabetes. ■

Erika Check

1. Baur, J. A. et al. *Nature* doi:10.1038/nature05354 (2006).
2. Kaeberlein, M. & Rabinovitch, P. S. *Nature* doi:10.1038/nature05308 (2006).

Hwang takes the stand at fraud trial

SEOUL

Woo Suk Hwang has taken the stand in court in his defence for the first time. The once-lauded cloning expert was confident and at times defiant as he struck back at the prosecution's claims that he embezzled money, committed fraud and broke a bioethics law, at one point citing dealings with the Russian mafia to explain himself.

Hwang electrified the scientific world in 2004 and 2005 with announcements that he had cloned human embryos and extracted stem cells with great therapeutic potential. But all the results were then shown to be fake. Hwang was charged in May with fraud, embezzlement of 2.8 billion won (US\$3 million) and violation of bioethics legislation that outlaws the purchase of eggs for research. Several of his colleagues were also charged with various offences.

Tracking Hwang's research funding has not been easy, and it is still not clear where all the money came from or where it ended up. For example, receipts for hundreds of millions of won for buying pigs and cows were faked.



Woo Suk Hwang is in confident mood, despite serious allegations.

Prosecutors have struggled to track Hwang's expenses through at least 63 bank accounts carrying the names of members of his research team, a high-school classmate and a relative, as well as pig and cow dealers. Hwang first denied but later admitted using these accounts. According to the prosecutor's report, Hwang used to carry a bag full of cash to different bank branches, making relatively small deposits in various accounts minutes or hours apart.

The defence denies embezzlement, claiming that Hwang was not trying to fill his own coffers. Hwang's lawyers explained that money earmarked for buying cows and paying for junior lab members' housing was simply spent as part of overall research expenditures. "Hwang had no intention to make himself rich," says Keun-Hwa Jung, one of a team of at least five independent lawyers working on the case, pointing out that a tally of research spending shows that Hwang spent more than he received.

When he got the chance to put his side of the story, Hwang strode into court looking relaxed,



shaking hands with supporters. He explained some of the faked receipts by saying that he had given money to a Russian mafia group, as part of an effort to get tissue samples of a mammoth he was hoping to clone. He could not get receipts for those funds, Hwang told the court. The story is not as far-fetched as it might seem: there are ongoing mammoth-cloning projects in Japan. Hwang was keen to clone many different animals, including an endangered Korean tiger. And there is a black market for mammoth remains. Still, it is not clear how Hwang would have explained where he got the tissues if he had succeeded in creating a mammoth.

The lawyers also fended off the fraud allega-

Safer embryo tests could boost IVF pregnancy rates

NEW ORLEANS

It may soon be possible to assess the health of an embryo created by *in vitro* fertilization (IVF) without interfering with a single cell on its future body. New assays that look for tell-tale molecules secreted by healthy embryos could help reproductive biologists select the embryos that are most likely to produce a baby.

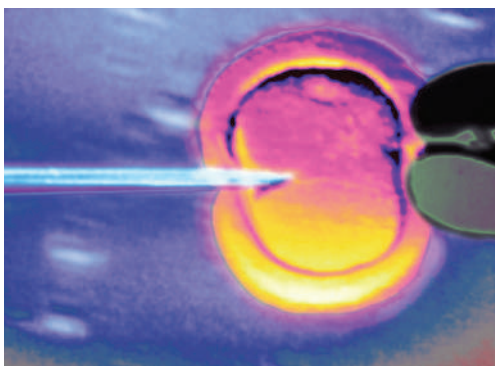
In IVF, several embryos are usually created. Embryologists try to select the healthiest by visual inspection of the number of cells or the rate at which they divide, and transfer these back to the womb.

But these assessments are subjective and fallible. In the

United States, on average less than 40% of IVF cycles end in pregnancy. To improve the odds, doctors routinely transfer multiple embryos, increasing the risk of twins or triplets.

Now researchers are working on selection methods that are quicker, more objective and leave the embryo intact. Several research groups and companies presented preliminary studies at the meeting of the American Society for Reproductive Medicine in New Orleans last week.

Such tests could boost pregnancy rates as well as reduce the need to transfer several embryos and risk multiple births.



Select few: analysis of proteins secreted by embryos could provide more accurate healthy embryo selection in IVF treatment.

"Everyone's arriving at the same realization," says Dagan Wells, who studies gene expression in eggs at Yale University in New Haven, Connecticut.

Many groups are trying to

find a fingerprint of proteins or metabolites that only healthy embryos secrete. Emre Seli at Yale and David Burns at McGill University in Montreal, Canada, have used spectroscopy to identify



Courting support: Hwang's testimony struck back at prosecutors in front of a courtroom of supporters.

tions. The prosecution says that Hwang secured billions of Korean won from private companies using data that he knew to be faked. But Hwang claims he believed that at least some of the cell lines, including the original cell line presented in 2004, were real. Hwang admits ordering the fabrication of certain images, including slides of teratomas. The ability of stem-cell lines to form teratomas proves their ability to form virtually all types of body cells. But the tests take months, so when they did not work out, Hwang says he ordered faked photographs "to save time". He maintains, however, contrary to

testimony by other researchers on the team, that he was not involved in any tests that would indicate he knew that the cell lines were fake, such as DNA fingerprinting tests that are necessary to match cell lines to the donor from which they are cloned.

Hwang's lawyers already have their case prepared for the next hearing, which will deal with the allegation that Hwang violated the country's bioethics law when he used eggs from women who were given free IVF treatments in exchange for donating eggs. The law prohibits the provision of any "financial reward,

property, or any other personal benefits" for the donation of sperm or eggs. But the lawyers will claim that illegal benefits do not include IVF treatment. "We have to rely on the spirit of the law here because the law itself is very unclear," says Jung.

Jung says that he hopes the judge will hand down a verdict by the end of the year. Hwang seems confident, and in anticipation of his acquittal has started up a new research institute in the outskirts of Seoul. One researcher there, who did not wish to be named without Hwang's approval, said he is very happy to be able to work with Hwang, which would have been unthinkable before. The institute is reportedly trying to clone pigs with humanized organs for transplantation, but the researcher refused to give any details.

If that happens, Hwang is likely to find support for his work. Some two hundred supporters attended the court to listen to Hwang's words, often nodding in solemn agreement. One was Zoon-hwan Go, a professor at Kyonggi University and author of a book, *Entrapment*, about a conspiracy against Hwang. "Hwang is diligent, truthful and pure," he told *Nature* during the trial. "I don't think he is perfectly clean," said another attendee. "But now everyone just thinks he's a liar. That's not right." She would be willing, she says, to donate eggs if Hwang were to start human research again.

Outside the courtroom, however, the country has mostly lost interest in the case. Many Koreans say it is an embarrassment they would like to forget.

David Cyranoski



WOO SUK HWANG

Find all of *Nature's* material on this case, complete with timeline, a guide to who's who, and the latest news. www.nature.com/news

YOU SUNG-HO/REUTERS

such a signature in the media in which embryos are grown. Two trials reported at the meeting showed that the test could predict which embryos would implant with more than 70% accuracy. Seli says he will start recruiting women into a larger clinical trial within months.

Others are trying to assess eggs after they are collected but before fertilization. Selecting the most promising could allow fewer embryos to be created. This would be particularly useful in countries with legal restrictions such as Italy, where the law permits only three eggs at a time to be fertilized.

A group led by Samir Hamamah, who heads the assisted-reproduction unit at the Hôpital Arnaud de Villeneuve in Montpellier, France, is extracting

and analysing the proteins in the cumulus cells that nestle intimately around an egg when it is ovulated. Hamamah reported in New Orleans that he has identified a handful of proteins that are manufactured at higher or lower levels in the cumulus cells of human eggs that go on to be fertilized compared with those that do not (S. Hamamah *et al.* *RBM Online*, www.rbmonline.com/Article/2559).

Like any development in reproductive biology, such tests are unlikely to come without controversy. Some researchers claim that tests for an antigen called HLA-G in the growth medium of an embryo can predict the outcome of

"Everyone's arriving at the same realization."

IVF. But not everyone is convinced that its levels really correlate with pregnancy. Any new test should be shown to boost pregnancy rates in clinical trials before it is offered to patients, says Mandy Katz-Jaffe at the Colorado Center for Reproductive Medicine in Englewood.

Katz-Jaffe believes she can take such screening methods one step further. She is working on an ambitious test that will not only distinguish healthy embryos, but reveal whether or not an embryo carries a specific chromosomal abnormality.

At present, technicians can extract a cell from an embryo and test it for genetic defects, a

procedure called preimplantation genetic diagnosis (PGD). But it's unclear whether this subtly damages the embryo. Katz-Jaffe is searching for proteins that are secreted at different levels by embryos with specific chromosomal defects, for example an extra copy of chromosome 21.

But Santigao Munne, an expert in PGD at diagnostic genetics company Reprogenetics in Livingston, New Jersey, questions whether this approach can replace PGD: "It's a lot of noise," he says. He believes that rather than detecting signs of particular abnormalities, such assays may simply be picking up molecules released by embryos that have stopped growing.

Helen Pearson

German science academy stumbles at birth

Power brokers in Germany's academic politics unexpectedly failed last week to approve the creation of a German Academy of Sciences.

An academy that could speak for the whole of German science has been under discussion since reunification in 1990 (see *Nature* 443, 371–372; 2006). The heads of Germany's main research organizations had agreed on a detailed concept for such a body, which would also provide scientific advice for politicians. But final talks between the research ministers of Germany's 16 states and the federal government stalled when Peter Gruss, the president of the Max Planck Society, wrote a last-minute protest letter saying he feared that his society would lose decision-making power internationally.

The ministers also failed to agree on how the costs of the proposed academy should be shared between them. They may reconsider the proposal next spring.

Los Alamos computer disk was 'traded for meth'

Classified data from Los Alamos National Laboratory, the first US nuclear-weapons lab, turned up last week at the home of a drug dealer in New Mexico.

The Federal Bureau of Investigation is investigating the breach, the latest in a string of security-related embarrassments for the lab. In 2000, two hard drives containing classified data disappeared briefly. And in 2004, the lab was shut down for months after two disks were reported missing; it later turned out that the disks never existed (see *Nature* 433, 447; 2005).

Local police have now found three flash drives containing classified material during a search of a trailer home occupied by a known methamphetamine dealer. A former subcontractor at the lab was also living there. At least one of the drives was traded for methamphetamine, the dealer told the local newspaper from his jail cell.

Los Alamos director Michael Anastasio said the lab is trying to find out what went wrong. "This is a serious matter, and we are taking immediate steps to address it," he says.

NIH researchers fear ethics rules will hit recruitment

The ethics rules drawn up by the US National Institutes of Health may look good to the public, but employees are still worried that they may cause staff to leave or deter researchers from joining the agency. Those

Sweet success shows you can count on the public

WDR

Are the masses stupid? A hundred years ago Francis Galton, a half-cousin of Charles Darwin, analysed the results of a guess-the-weight-of-an-ox competition at a country fair. The average of the guesses from 787 participants was virtually spot-on to the real weight of the ox, Galton reported (see *Nature* **75**, 450; 1907). The findings shook his belief in eugenics, a term he himself had coined.

Last month, Ranga Yogeshwar, host of Germany's popular *Quarks & Co* science television show, repeated the experiment — but with a laboratory beaker filled with sweets. The average guess of the 16,000 viewers was again astonishingly close to the correct number of 5,780; the average was 5,714, and nearly one in 200 hit the right number exactly. "The masses are intelligent," says Yogeshwar. "We can use them for all sorts of science TV-Internet experiments."



are just two of the messages emerging from an independent survey commissioned by the NIH and released on 26 October, to assess the impact of the regulations.

In August 2005, after scandals over conflicts of interest, the NIH announced more stringent rules, including a ban on outside consulting. The Internet survey, carried out this summer, received responses from some 8,000 agency employees (a 48% response), including around 3,300 scientists.

Nearly three-quarters of employees thought the rules would improve the NIH's credibility with the public, but more than half thought they would harm the agency's ability to retain or recruit employees. Among tenure and tenure-track researchers, 18% said they were looking to leave the agency, or considering doing so. Yet nearly 90% of all agency scientists said they planned to still be at the agency in a year's time.

The NIH will next survey employees who left the agency, along with potential new employees, to see the effect of the ethics rules on their decisions.

South Korea finds time and cash for stem cells

There's hope — and money — in South Korea for stem-cell researchers wanting to move on from the Woo Suk Hwang fraud scandal (see page 12). To get the word out, scientists are hosting a handful of one-day international symposia, including one on 3 November in Seoul on mesenchymal stem cells, which are found in bone marrow.

The goal is to stimulate plans for the 430 billion won (US\$450 million) that the South Korean government promised in May for research over the next 10 years; 30% of the money will be put towards research on embryonic stem cells, with the

rest going on adult stem-cell research.

Most of the embryo work will be led by the Stem Cell Research Center based in Seoul, which gets some \$15 million a year until 2012. Another initiative, launched by the science ministry in August, will provide \$4.5 million a year in grants for groups working on stem-cell differentiation and other related studies.

DNA catalogue opens up new era of mouse research

Mouse researchers, rejoice: scientists have completed a two-year, \$13-million project that should catapult mouse studies into a new era.

The project identified 8.3 million differences in DNA sequences — single nucleotide polymorphisms, or SNPs — from 15 mouse strains. This SNP catalogue should enable studies among several different strains of mice, rather than single inbred strains. And it will allow researchers to examine the combinations of environmental and genetic factors that influence the risk of diseases and the ill effects of toxins and chemicals.

The study, which was funded by the US National Institute of Environmental Health Sciences, was conducted by Perlegen Sciences of Mountain View, California. All the resulting data have been made freely available.

▶ www.ncbi.nlm.nih.gov/SNP



Correction

The story 'Funding agencies toughen stance on open access' (*Nature* **443**, 894–895; 2006) incorrectly stated that the American Physiological Society (APS) refused to comply with Wellcome Trust guidelines for making publications openly accessible. In fact, the APS is still in negotiations with the trust, and has not reached a final decision.

BUSINESS

Drilling for nanotech gold

One US nanotechnology start-up has hit the jackpot — but for others the prospect of such overnight success seems remote. **Colin Macilwain** reports.

The guys in the aisle at Home Depot don't know it. But that \$800 DeWalt cordless power-tool set — the one they really want for Christmas, but are just too scared to ask for — gets its butt-kicking oomph from a *Nature Materials* paper published only four years ago.

It's taken that time for a battery cathode based on phosphate nanocrystals to rip its way from a lab at the Massachusetts Institute of Technology (MIT) in Cambridge, through financing, design, development and manufacture in east Asia, to its current position, driving 36-volt power tools from Black & Decker — owner of the DeWalt professional-grade marque.

"These tools are better than corded ones," says Jamie Mann, director of sourcing for Black & Decker in North America. "And they can take 2,000 recharges. That's big for us — we think it's changed the game."

The company that builds the nanophosphate batteries, A123Systems in Watertown, Massachusetts, was founded in 2002 and has US\$100 million worth of orders in hand. It's an instant success story, and one that impressed investors gathered in Cambridge, Massachusetts, last week for the Lux Research Executive Summit, a meeting of the great and the would-be great in the fledgling nanotech sector.

But the MIT spin-off's success won't be matched by most of the estimated 1,500 start-up companies in the sector, the meeting was told. "In every new industry, you have an initial hype, and then a shake-out," says Charles Seeney, president of NanoBioMagnetics, an Oklahoma-based health-products company. "The same thing is going to happen in nanotechnology."

For those that do survive, partnerships with major corporations — like A123Systems' with Black & Decker — will probably hold the key.

Not that A123's founders — materials scientists Yet-Ming Chiang and Bart Riley, and Ric Fulop, an entrepreneur and business fellow at MIT — had much idea initially where their dramatic success was going to come from. They had solid hopes for the company, and attracted an accomplished business manager — David Vieau, a mechanical engineer with extensive high-technology experience — to come and run it. But the power-tool breakthrough was unexpected: "We were considering things with



motors — we thought a lot about hybrid vehicles," recalls Vieau. "We didn't know where our optimum performance advantage would lie."

The *Nature Materials* paper outlined the basic idea, showing how tiny lithium iron phosphate crystals could be 'doped' to conduct electricity, and proposing that they be used to make battery electrodes (S.-Y. Chung, J. T. Bloking & Y.-M. Chiang *Nature Mater.* **1**, 123–128; 2002). Conventional lithium-ion batteries use particles of lithium cobalt oxide about a micrometre across: if they were smaller their electrical conductivity would improve but so would their thermal conductivity — raising possible safety issues. The new material is more chemically stable, and its performance can be optimized by using crystals that are only a few nanometres across. The result prompted fierce controversy over the mechanism that

lay behind it (*Nature Mater.* **2**, 702–703; 2003).

Chiang says the idea for the venture came when Fulop had walked into his office, and said: "How about starting a battery company?"

The company didn't take long to get rolling. Chiang felt that it would be a far easier sell than the high-temperature superconducting wire made by his previous start-up, American Superconductor. There, he recalls, he was selling mainly to "utilities that run on 25-year-old technology". Whereas with the new business, "it is one of the easiest things in the world to explain why we need better batteries". The company was taking shape before the paper was published: it hired Vieau in March 2002 and quickly obtained backing from a loyal clutch of

"It is one of the easiest things in the world to explain why we need better batteries."

— Yet-Ming Chiang

Big boys' toys: but there's nanotechnology inside.

investors, including North Bridge Venture Partners of Boston, Sequoia Capital of Menlo Park, California, and investor Desh Deshpande, who became company chairman.

At the time, Black & Decker was searching for a breakthrough that would get it a jump ahead in the fast-expanding cordless power-tool business. "We were scouring the Earth," says Mann. After meeting with the young company, "we were excited, but extremely cautious," he recalls. "We're a \$5-billion company, and this was ten guys from Boston! It looked like a good bet, but there was a lot of risk."

The odds were lengthened by the nature of the lithium-ion battery business, which is dominated by huge Japanese corporations, such as Sony, with deep pockets and large, intramural

research efforts of their own. Patent lawyers trawled to see if these companies were onto the phosphates breakthrough. But they weren't.

Mann told Vieau that, if they could succeed in making the powder in bulk, there was potential for a deal. By early 2004, A123Systems could make 1 kg of the compound a month — with a manufacturer in Taiwan that promptly went bust. The company raised a further \$20 million to keep itself going, made its first battery prototypes, and in May 2005 signed an agreement with Black & Decker.

This agreement precludes A123Systems from selling the technology to Black & Decker's rivals while assuring volume sales for its batteries. It's a type of arrangement that is increasingly fundamental to the prospects of nanotechnology start-ups. Black & Decker raised the idea of taking an equity stake, but the small company demurred (Motorola, the electronics company, already has a shareholding). A public share offering is being considered, Vieau says, but is at least 12 months away.

This runaway success is, to put it mildly, typical of a nanotech start-up. Lux Research, the New York consultancy that organized the Cambridge meeting, has just tried to grade the attractiveness as corporate partners of 136 companies that have venture-capital backing. A123Systems tied second with Aspen Aerogels, an insulation company based in Northborough, Massachusetts. NanoOpto, an optical-equipment supplier in Somerset, New Jersey, came out top.

The fashionable nanotech label doesn't always help in impressing investors. The phrase "is a fantastic indicator of where innovation is coming from," says Greg Schmergel, president of memory-chip firm Nantero in Woburn, Massachusetts. But when it gets down to nitty-gritty, he quips, "I've always found that being a nanotechnology company is a great help — for about the first minute of a meeting with investors."

But supporters maintain that nanotech is coming of age. "Nanotechnology's time is now," claims Lux Research president Matthew Nordan, who says the field is at "an inflexion point", after which "discovery continues, but it gets eclipsed by commercialization".

The chemistry of a deal with a larger partner matters just as much as the chemistry in the lab. As Mann and Vieau run through their partnership's short but lively history, the mutual respect and understanding between the two is palpable.

But the real key to success is to provide a solution to a real need, whether it's for health-care or for electric drills. The best-equipped outfits, according to Ken Morse, director of the MIT Entrepreneurship Center, "are not in the nanotech business — they are in the solutions business."

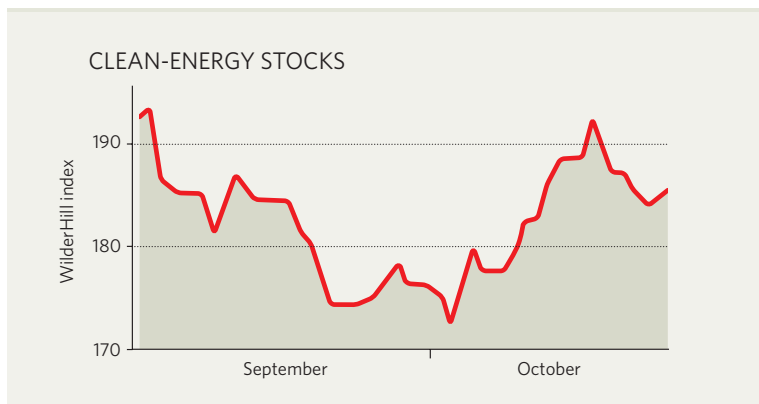
IN BRIEF

BATTERY BREAKDOWN Sony announced operating losses for the second quarter as the consumer electronics giant began to absorb the cost of recalling several million of its lithium-ion batteries. The company announced an operating loss of ¥21 billion (US\$178 million) for the quarter ending 30 September, after taking a ¥51.2-billion charge for the recall, which has been undertaken because of a small risk that the batteries will catch fire. Analysts expressed concern that the recall is taking some of the shine off Sony's sharp image as a premium brand.

ACTIVIST REIGNS A frustrated investor has taken over as chairman of ImClone Systems, the one-time biotech star that has enjoyed a rocky ride since its founder, Samuel Waksal, was ousted for insider-trading in 2002. Carl Ichan, a billionaire with no biotechnology experience, took over on 24 October, and three board directors and the interim chief executive, Joseph Fischer, departed. The New York company's main product is the cancer drug Erbitux, which is facing competition from Vectibix, a rival compound made by Amgen. Ichan says he'll concentrate on getting ImClone's marketing partner, Bristol-Myers Squibb, to sell the drug more effectively.

NUCLEAR FALLOUT The UK government said it is abandoning plans to privatize the British Nuclear Group, its nuclear clean-up operation, as a going concern. The group is now likely to be broken up and sold in small parts, trade and industry secretary Alistair Darling said. The change of plan came when it emerged that likely bidders for the group were likely to break it up and sell off its components, according to government officials. These operations should flourish: the projected cost of cleaning up Britain's nuclear complex is now £65 billion (\$122 billion).

MARKET WATCH



Stocks in clean-energy companies held steady over the past two months — despite a backdrop of falling oil prices that, sceptics fear, could deflate the enthusiasm that investors felt for the sector earlier in the year.

The WilderHill Clean Energy index (symbol ECO on the American Stock Exchange) dipped in September but recovered last month. "It's become fairly non-correlated to the oil price — which I like," says Robert Wilder, founder of WilderShares, the California consultancy that compiles the index on the basis of the stock prices of suppliers of non-fossil, non-nuclear energy sources.

These suppliers are now hopeful that the drop in the price of a barrel of crude oil — from more than \$70 in August to \$55 last month — won't deflate the boom in demand for their products. "I see no major risks unless the oil price

goes below around \$45," says Michael Liebreich, chief executive of London-based New Energy Finance. "At that point some investors might lose their nerve and pull out of some projects. If that happens, stock prices will wobble."

After a summer in which several ambitious plans for stockmarket offerings were abandoned, German biofuels suppliers CropEnergies and Verbio managed to raise some €400 million (\$500 million) between them on the Frankfurt stock exchange in October.

Liebreich draws encouragement from the relative stability of the market. "There was an anomalous period, from President George Bush's State of the Union speech on oil addiction until the market correction in May," he says, "but the market has reacted well and rationality has returned."

Colin Macilwain



The war in Iraq, the price of oil, the deadlock over Iran's nuclear ambitions, the terrorism of al-Qaeda and the tensions surrounding immigrant communities in Europe ensure that Islam is rarely far from the headlines. But you would have to be an avid student of Muslim affairs to come across any discussion of science and technology not linked to the development of nuclear weapons.

In this week's issue, *Nature* offers an unprecedented look at the prospects for science and technology in the Muslim world (see page 20). We have never before collected together such a range of voices and analysis in one issue.

In ignoring Muslim science, the West follows the lead of the Muslim world itself. Low investment and a low profile combine to keep the scientific community small, marginalized and unproductive. This is not simply a matter of underdevelopment; the oil-rich Gulf states invest pitifully in R&D (see page 28). In our Commentary section, on pages 33 and 35, Nader Fergany, the lead author of the Arab Human Development Reports, and Herwig Schopper, president of the council for the Middle East laboratory SESAME, offer their own critical analyses of what needs to change to allow science to take off in Muslim countries.

The poor scientific track record of Islamic countries might suggest that there is something about Islam inherently inimical to research. Muslims bristle at this idea, pointing to the major achievements of Muslim scholars under the Islamic caliphate (see timeline, page 23). But what of the present? Our News Feature on page 22 looks at the attitudes to science in the various Islamist organizations growing in power in key states ranging from the Occupied Palestinian Territory to Malaysia. The secular regimes and one-party states that have ruled many Muslim countries are being replaced, or directly challenged, by voices calling for a more directly political Islam.

The conditions in which knowledge flowered a millennium ago are hardly those that today's Islamists say they favour. Back then, support for scientific enquiry was matched by an openness to other cultures and sources of knowledge. But when Islamists come to power the picture is more nuanced than it may first appear. Restrictions on freedom of speech and a high level of investment in military technology are distressing to outsiders, but greater attention to higher education is a trend that could offer hope. Mostafa Moin, an Iranian reformer and scientist, lays out his hopes and fears for the future on page 29.

Greater attention to the challenges of the present is sorely needed. Too few Muslim governments collect data on the status of science and innovation (as our analysis on page 26 shows), and so the problems facing scientists are not even on their agenda. Muslim nations wanting to invest in science as a broad cultural activity need to extract the right lessons from their glorious past and their politically charged present. ■

N. ELAWADY

AMBITION & NEGLECT

Science in the Muslim world

The Islamic world

The 57 countries in the Organization of the Islamic Conference are home to 1.3 billion people. The attendant diversity in culture, geography, economics and politics can be seen in these snapshots of five different approaches to science.

IRAN

Revolution and reform

The decline of scientific knowledge in Persia, now Iran, began in the fifteenth century, as the entire Islamic world lost touch with its intellectual roots.

In the 1970s, the last Shah dynasty attempted to reverse the trend by building new universities and sending students abroad to do PhDs, after which many returned to teach and research in Iran. The number of scientific publications crept up from 125 in 1970 to nearly 400 in 1979, the year of the Islamic revolution.

The revolution halted this modest progress. Many of the scientific élite fled the country as the new regime closed universities and turned against 'Western science'. The Iran-Iraq war in the 1980s further drained resources, leaving little money for non-military education and research.

With peace, more public money became available and Islamic fervour moderated to accept the intrinsic value of science. Universities expanded again, and were

allowed to award PhDs. The publication rate also rose, more sharply than in the 1970s: by 2003 the annual output of papers reached close to 2,000.

With the election last year of President Mahmoud Ahmadinejad, a hardline Islamist, every aspect of Iranian life — from its supreme leader, to the judiciary, and parliament — is now in religiously conservative hands. Iran's reformers have gone underground.

Many areas of science continue to be well funded, including stem-cell research. Iran was the first Middle East country to develop a human embryonic stem-cell line, using spare embryos from *in vitro* fertilization. And despite increasing restrictions on free speech, higher education continues to expand. Whether the new university presidents — all appointed by Ahmadinejad — will yield to conservative demands for greater controls at universities remains to be seen.

Alison Abbott

INDONESIA

Boom and bust

Asked the greatest achievement of his institute in the past five years, Sangkot Marzuki, director of the Eijkman Institute for Molecular Biology in Jakarta, says: "Still being here."

The 1990s were a boom time for Indonesian science. In 1993, the Eijkman Institute was resurrected after political, social and economic unrest closed its doors in 1965. But the Asian financial crisis of 1997 threatened to reverse the gains.

In 1998, following two years of social unrest and demonstrations, Indonesia's autocratic leader, General Suharto, stepped down, ushering in free elections and reform. Since 2004, the president has been freely elected. Religion does not seem to have the same link with power as in other predominantly Muslim countries. Nearly 90% of Indonesians are Muslims, but neither of the main political parties is strongly linked with Islam. Still, opposition by many smaller Islamic parties was crucial in quashing the bid by a woman, Megawati Sukarnoputri, to retain her position as president in 2004.

Religious friction has hurt science by threatening international collaborations. Marzuki says that before 1998 his institute had six Australian scientists on three-year grants from the Australian government. "With the travel restrictions frequently issued, especially after the Bali bombing, no collaboration has been possible," he says.

Islam itself is very flexible in relation to science, says Marzuki: "Our institute performs prenatal diagnosis against common genetic diseases, in particular thalassaemia. The government is against reproductive cloning but supportive of therapeutic cloning, a position adopted on the recommendation of the Indonesian Academy of Sciences."

With stability, Indonesia could be a fertile ground for science. The government, when cash flow allows, supports research. But no one welcomes the misfortunes that have forced research in some fields — earthquakes, tsunami monitoring, avian influenza — on Indonesian scientists.

David Cyranoski

MALAYSIA

High-tech hopes and fears

Malaysia has one of the most technologically advanced societies in the world, thanks in part to its consumer-electronics industry. It outshines the rest of the Muslim world in high-tech exports, but is not resting on its laurels.

The government, led since the Malaysians received their independence in 1957 by the United Malays National Organisation, is keen to build on its successes in making semiconductor components and to develop new technologies. But government policies have driven away some of the country's best talent and kept Malaysia isolated from international science. Too often, pleasing the Islamic Malays, who make up the majority of the population, takes precedence over rewarding scientific merit.

Historic tensions between the Malays and the Chinese minority, who have long held the economic reins, explain some of these policies. In 1969, riots followed general elections in which Chinese parties made gains.

The government has faced a difficult balancing act since then.

The most controversial policy is a university quota system that favours ethnic Malays (some 60% of the population), over Chinese (25%) and Indian (7–10%). Minority students, many of whom have top grades, struggle to get into the nation's best universities, and often end up going to the United States, Singapore or Australia. Such policies also inhibit interactions with the international community. And attempts to reverse a mainly Chinese brain drain have failed.

The government has long invested in large projects intended to benefit high-tech industry, but with little success. And ongoing privatization of government operations in various sectors, including roads, energy and technology, is slanted to help the Malays. Such 'Malay first' policies will fail to attract the best overseas talent and continue to leave Malaysia isolated.

David Cyranoski

PAKISTAN

Tanks and technology

Pakistan's history of perpetual military coups is a disaster for democracy. But the men in tanks have been more generous than civilians when it comes to funding science and technology.

Pakistan's national biotechnology institute, for example, owes its existence to former President General Zia ul-Haq (military coup, 1977). In contrast, a decade of civilian rule after the general's death saw research funding drop to critically low levels — former Prime Minister Benazir Bhutto did not even appoint a science minister. Credit for current science and technology funding — the highest it has ever been — goes to President General Pervez Musharraf (military coup, 1999).

Musharraf has handed the task of reorganizing research and higher education to Atta-ur-Rahman, a chemistry professor at the University of Karachi. Rahman's many reforms include increasing the number of universities and sending more students

abroad to train. Schemes to attract foreign faculty members to work in Pakistan's universities, and performance-related pay for the country's own academics, have been more controversial.

Critics of the reforms include Pervez Hoodbhoy, physics professor at Quaid-i-Azam University in Islamabad. He says that formerly cash-starved ministries lack the management capacity to spend their windfall wisely, and that the expanding university numbers have not been matched by a commitment to quality. Pakistani researchers, meanwhile, are unhappy that foreign faculty members get higher salaries for the same work.

Rahman says he understands the criticisms. But he is a man in a hurry. General Zia's life ended in a plane crash in 1988. There have already been two attempts on Musharraf's life. Rahman knows the boom will end when the general leaves office.

Ehsan Masood

SAUDI ARABIA

Slow starter

A year after the birth of independent Saudi Arabia in 1932, the kingdom's rulers faced a problem with the public acceptance of technology. The dilemma was caused when US oil companies prospecting in the region requested permission to take aerial images of the desert.

Newly crowned King Ibn Saud, founder of the present Saudi dynasty, had two concerns. First, that indigenous tribes might shoot at something they considered to be extraterrestrial. Second, if religious authorities believed that the onboard cameras could glimpse the face of God, it would invite divine wrath.

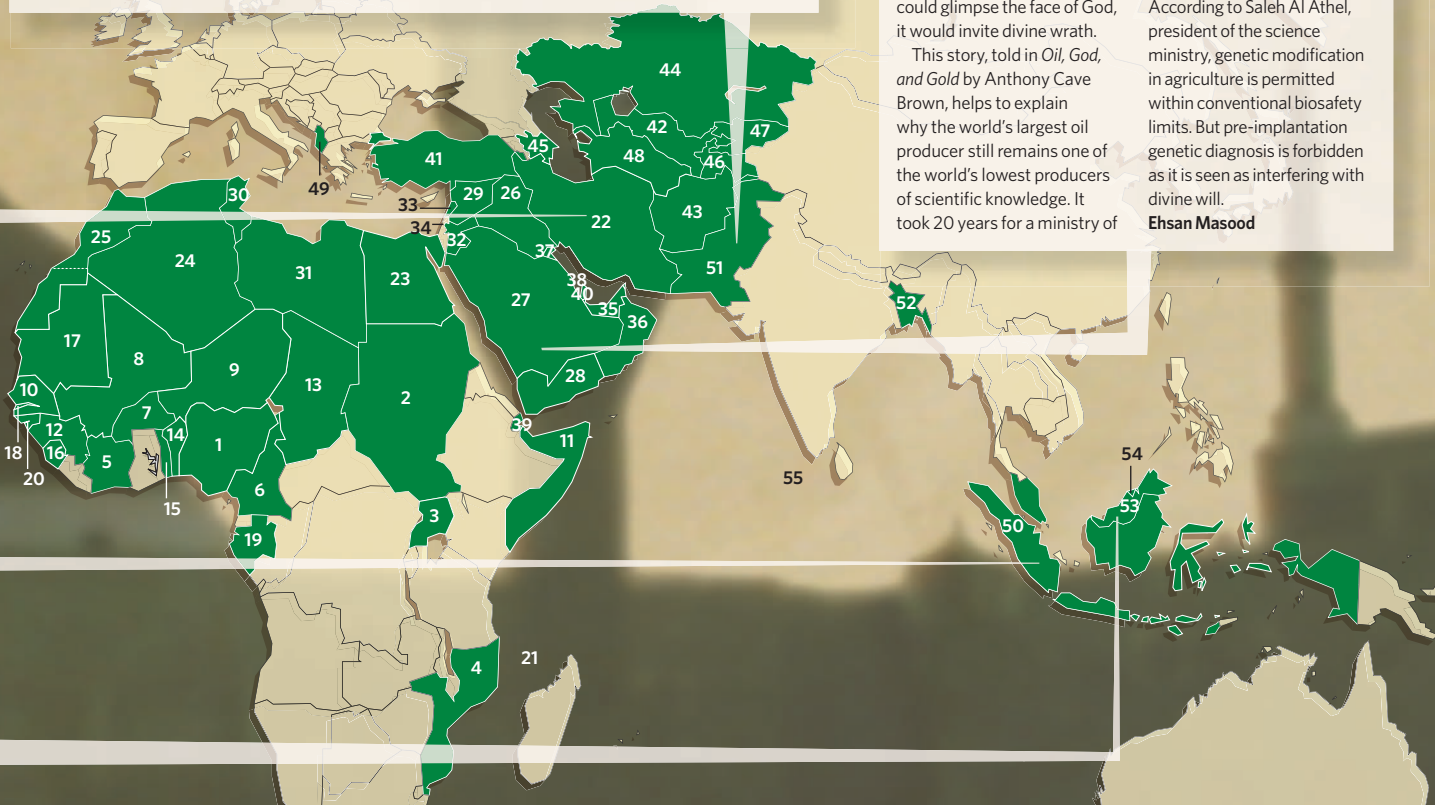
This story, told in *Oil, God, and Gold* by Anthony Cave Brown, helps to explain why the world's largest oil producer still remains one of the world's lowest producers of scientific knowledge. It took 20 years for a ministry of

education to be created. The country's science ministry did not emerge until 1977. Even today, Saudi Arabia spends just 0.25% of its gross domestic product on science and technology.

Despite this, there are some bright spots. The number of specialist science and engineering colleges has doubled to 64 in the past decade, and the number of students enrolled in related degrees has also doubled to 76,000.

Moreover, Saudi society has progressed considerably in its acceptance of new technology. According to Saleh Al Athel, president of the science ministry, genetic modification in agriculture is permitted within conventional biosafety limits. But pre-implantation genetic diagnosis is forbidden as it is seen as interfering with divine will.

Ehsan Masood



POPULATION OF THE OIC COUNTRIES (total, thousands, in 2000)*

Sub-Saharan Africa

1. Nigeria	113,862
2. Sudan	31,095
3. Uganda	23,300
4. Mozambique	18,292
5. Côte d'Ivoire	16,013
6. Cameroon	14,876
7. Burkina Faso	11,535
8. Mali	11,351
9. Niger	10,832
10. Senegal	9,421
11. Somalia	8,778
12. Guinea	8,154
13. Chad	7,885
14. Benin	6,272
15. Togo	4,527
16. Sierra Leone	4,405

17. Mauritania	2,665
18. Gambia	1,303
19. Gabon	1,230
20. Guinea-Bissau	1,199
21. Comoros	706

Middle East & North Africa

22. Iran	70,330
23. Egypt	67,884
24. Algeria	30,291
25. Morocco (including Western Sahara)	29,878
26. Iraq	22,946
27. Saudi Arabia	20,346
28. Yemen	18,349
29. Syria	16,189
30. Tunisia	9,459

31. Libya	5,290
32. Jordan	4,913
33. Lebanon	3,496
34. Occupied Palestinian Territory	3,191
35. United Arab Emirates	2,606
36. Oman	2,538
37. Kuwait	1,914
38. Bahrain	640
39. Djibouti	632
40. Qatar	565

Europe & Central Asia

41. Turkey	66,668
42. Uzbekistan	24,881
43. Afghanistan	21,765
44. Kazakhstan	16,172

45. Azerbaijan	8,041
46. Tajikistan	6,087
47. Kyrgyzstan	4,921
48. Turkmenistan	4,737
49. Albania	3,134

South & East Asia and Pacific

50. Indonesia	212,092
51. Pakistan	141,256
52. Bangladesh	137,439
53. Malaysia	22,218
54. Brunei Darussalam	328
55. Maldives	291

Latin America & Caribbean

56. Guyana	761
57. Suriname	417

*Many countries outside the OIC have sizeable Muslim populations, notably India, which is home to 175 million Muslims.



An Islamist revolution

Islamist political parties are taking over from secular ones across the Muslim world. What does this mean for science at home and scientific cooperation with the West? **Ehsan Masood** investigates.

At Peshawar University on the Grand Trunk Road linking Pakistan, India and Bangladesh, there is much talk of growth. Its national centre for excellence in geology is to get 11 new labs, a library and a new museum. The provincial government, moreover, has handed the university the job of running a botanical garden and a 40.5-hectare national park.

Peshawar is the capital city of Pakistan's north-west frontier province, the border region with Afghanistan where the Taliban first emerged among the Afghan refugee population in the 1990s. None of the university's activities is unusual for a leading institute in a developing country. But what might seem surprising to outsiders is that, after many years of neglect, the university's expansion comes at a time when local people have elected an alliance of political parties which, like the Taliban, want to base most laws on the Koran. Unusually for Pakistan, the current provincial government has forbidden male doctors from attending to female patients and has banned music on public transport.

The university is run by Haroon Rashid, a professor of chemistry who was appointed vice-chancellor in January 2006. In common with the majority of Pakistanis, Rashid is a Muslim,

something that he is proud to make known. Could a university vice-chancellor in Peshawar be of any other faith? In today's Peshawar, a non-Muslim vice-chancellor would be next to impossible.

Pakistan, along with the Islamic Republic of Iran and Sudan, has been run by governments that put Islam at the centre of politics for many years. As more Muslim countries give their citizens the right to vote, Islamist political groupings have taken power, or form the main opposition, in national or regional assemblies in Iraq, Kuwait, the Occupied Palestinian Territory, Bahrain, Egypt, Afghanistan, Jordan, Morocco, Malaysia and Turkey. Islamist is a term used to denote those committed to the application of Islamic principles and Islamic law in politics.

What can Muslim scientists expect from the new Islamist parties that are seeking power across the Muslim world? Will there be more support for science and for research infrastructure, as in Peshawar, but an environment where basic freedoms continue to be denied? The mostly secular, although undemocratic, regimes that have hitherto ruled for decades across the Muslim world have rarely paid more than lip-service to investment in science and technology. Consequently, today's Muslim

states barely register on indices of research spending, patents and publications, and only Turkey has universities in the global top 500.

Much of this is candidly documented in the four volumes so far of the *Arab Human Development Report* from the United Nations Development Programme, written entirely by Arabic-speaking social and natural scientists (see page 33), which lays bare how knowledge-based activities such as science, innovation, book publishing, art and literature in Arabic-speaking countries are among the weakest in the world. The report does not consider non-Arab member states of the 57-strong Organization of the Islamic Conference (OIC), such as Indonesia, Pakistan and Turkey. But, as the data on page 26 show, the picture in the broader Muslim world is not much better.

The situation for Muslim science has been bad, and one assumption, based on current trends, is that things can only get worse. One fear is further restrictions on freedom of expression. Political leaders in the Muslim world, even in countries run on strict secular lines, are famously intolerant of dissent, as last year's attempted prosecution in Turkey of Orhan Pamuk, this year's winner of the Nobel prize for literature, demonstrates. Pamuk was accused of

N. NOUR/AP

insulting Turkishness. Even today, few universities enjoy much autonomy, and appointments to research posts are opaque and prone to corruption. If secular governments did little for science, can Islamist ones be any worse?

In the search for answers, Egypt's Muslim Brotherhood is a good place to start. The grandparent of Islamists, the brotherhood is a political party founded in Egypt in 1928. Its original aims included taking power, opposing Western influence in Egyptian politics, and governing using the Koran as the basis for lawmaking.

Mixed message

The party's presence and influence has expanded across the Muslim world — from the Middle East to Africa and Asia. In the absence of basic infrastructure in many countries, the brotherhood and its sister organizations run schools and hospitals, and its members include many scientists. But officially it does not exist — it is banned everywhere, and membership can be punishable by long spells in prison. To avoid censure its members stand as independents at election time, or as members of alternative parties. In Egypt, 88 brotherhood members of parliament together form the largest grouping after that of the government.

Kamal El Helbawi, who now lives in London, is a one-time senior official in the Muslim Brotherhood, and its former spokesman in Europe. In common with, arguably, most Muslims, Helbawi sees science and Islam as being in harmony, and he says that any government led by the Muslim Brotherhood will reverse decades of underinvestment in R&D. Is this a rose-tinted view or a genuine commitment? The answer may depend on the resonance of science and technology with the wider debates occurring in Muslim society. It may also depend on whether Islamist parties lean towards the Shia or Sunni schools of thinking (see 'A long tradition', page 24).

For Helbawi, science has three functions in society. First, it is a set of tools to help humankind enjoy a higher quality of life through new technologies or by solving problems that afflict the poor. Second, science and technology can be used to deter aggression, a justification, Helbawi believes, for developing a nuclear deterrent. And third, Helbawi believes that science has a role in strengthening religious belief. In his view, the Koran, in addition to being the word of God, was designed by God to convince

doubters of the truth of Islam and of creation. "I urge all scientists to read the Koran, from which they will learn much about so many scientific topics," he says.

Like many Islamists, Helbawi peppers his explanations with quotes from the Koran. He does so to underline that these are not his opinions — they have divine endorsement. For example, in explaining support for a nuclear deterrent he quotes chapter 8, verse 60. "Hence make ready against them whatever force and war mounts you are able to muster, so that you might deter thereby the enemies of God."

Listening to Helbawi, it seems that although science investment may go up, the space to disagree with the official line will go down. Yet within the brotherhood itself, there is much debate on literalism, reason and rationality, suggesting that totalitarianism is not the only option. Among the rationalists, for example, is Tariq Ramadan, a philosopher of religion at the University of Oxford and the maternal grandson of Hassan Al Banna, the Muslim Brotherhood's founder. Ramadan says that the Koran should not be quoted outside of its religious and historical context. He also worries that Helbawi's literalism amounts to an invitation not to think, and to assume, for example, that if all science is contained in the Koran, there is no place in society for new knowledge.

For Muslim societies, a literal interpretation of the Koran would present as many barriers to science and to freedom of thought as did the secular governments of the past. But the picture becomes more nuanced the closer one looks at Islamist governments once they are in power. Using Sudan, Pakistan and Iran as examples of countries where Islam is prominent in politics and which may foreshadow what may follow elsewhere, certain trends are clear.

In the case of Iran and Pakistan, there has been a substantial expansion in higher education and more spending on research, measures to improve scientific quality, and some opening up of labs to scientists from overseas. Iran's university population has swelled from 100,000 in 1979 to 2 million today. Pakistan's university population has increased from 276,000 in 2001 to 423,000 in 2004. Sudan's public-sector universities, too, increased from 5 in 1989 to 26 in 1996. In each country, there are equal numbers of women and men entering many



"I urge all scientists to read the Koran, from which they will learn much about so many scientific topics."

— Kamal El Helbawi

ISLAMIC ERA SCIENCE

750-1258 ABBASID ERA

The Abbasids overthrew the Umayyad caliphate, which had spread Islam through Asia, the Middle East, north Africa and the Iberian peninsula. The Abbasids moved the caliphate's capital from Damascus to Baghdad. This was a particularly productive period for science in Islamic history.

c.721-c.813 Jabir ibn Hayyan

Works attributed to this alchemist had lasting influence in Europe until the sixteenth century. Many words in chemistry have Arabic roots including alkali (*al-qaliy*) and alcohol (*al-kohl*).

c.830-1258 House of Wisdom, Baghdad

Activities at this library and research centre included translation of Greek works into Arabic by both Muslim and non-Muslim scholars. Free public libraries later spread to other cities.



c.780-c.840 Al-Khwarizmi

Mathematician who gave his name to 'algorithm'. Latin translations of his books introduced algebra (derived from *al-jabr*) to Europe.

c.865-c.926 Al-Razi (Rhazes)

Persian who contributed to medicine, alchemy and philosophy. He formulated the first known description of smallpox, which the Ancient Greeks had confused with measles.

c.965-c.1039 Ibn al-Haytham (Alhazen)



Basra-born researcher in astronomy, mathematics and optics who helped to develop the camera obscura. He refuted Greek models of vision, arguing instead that vision is the result of images being formed, and provided better descriptions of the eye.

980-1037 Ibn Sina (Avicenna)

Persian physician and philosopher from Bukhara. The Latin translation of his *Al-Qanun fi al-Tibb* (*The Canon of Medicine*) was a highly regarded medical text in Europe until the sixteenth century.

M. NIKOUBAZLPH/TH/REUTERS

faculties. Indeed, in Iran some 70% of science and engineering students are women. This university expansion is, however, creating its own tensions as the economies are not large enough to absorb so many new graduates, particularly women.

Call to arms

Second, each country has directed funds towards military R&D, money that could, for example, have been spent on R&D towards alleviating poverty. Why the neglect of the poor? For many Islamists, achieving independence from Western nations, defence and national security are higher priorities than the Islamic duty to care for society's poorest. Iran, like Pakistan, insists on maintaining a capability to enrich uranium to weapons grade. Egypt and Turkey also both recently announced plans to develop nuclear power. Abdul Qadeer Khan, former director of Pakistan's nuclear programme, was a keen proponent of spreading nuclear technology to other Muslim nations. He is now under house arrest in Islamabad for selling uranium-enrichment technology to Iran, Libya and North Korea.

A third trend suggests that Islamist governments are likely to restrict academic freedoms as much (if not more) than the secular regimes they want to replace. Saudi Arabia, Sudan, Iran and Pakistan are very restrictive environments for certain kinds of researchers, especially social scientists, to work in. Research into the role of government in public life, for example, requires governments to open up to the research community — something that these countries do not do. Because of this, the field of science and technology policy in all four countries is weak or non-existent.

Although academic freedom continues to be



In Iran, and elsewhere, equal numbers of men and women are studying science and engineering.

limited in Muslim countries, the field of Islamic theology is rife with debate and disagreement on many science-related topics. Moreover, thanks to cable television (in particular the Al Jazeera channel based in Qatar) and the Internet, this debate is beginning to be seen in public as never before. One keenly contested area for theologians is that of the ethics of new technologies. Another is evolution. Islamic opinion on bioethics varies widely, and different countries regulate in different ways. But on this issue, as others, public debate is not as free as it is in more open societies. Although theologians and scholars of religion debate among themselves, it needs a brave lay person or scientist (who is also conversant with theology) to challenge them in public.

Where do the differences in opinion lie? Saudi Arabia (an Islamic monarchy) and Iran, for example, have very different ideas on medical ethics. Saudi Arabia bans third-party *in vitro* fertilization on the grounds that sex and procreation is limited to husbands and wives. But third-party sperm donors are allowed in Iran because the alternative (a couple splitting up if they cannot have children) is considered worse for society. Similarly, Pakistan is practically alone in the Muslim world in banning organ donations from cadavers. This is because the country's Islamic authorities view the human body as being on loan from God, and when a person dies, the body needs to be returned to its creator close to its original state. But this view is not shared by other Muslim states.

Freedom to think

How literally they interpret the Koran will clearly influence how the new Islamist governments regulate science and technology. One of the Muslim Brotherhood's leading thinkers, the Egyptian scholar Yusuf Al-Qaradawi, who now lives in Qatar, is controversial in the West, but has mass support in the Arabic-speaking world, as well as among Muslims in Europe and North America. His book *Priorities of the Islamic Movement in the Coming Phase* (Awakening Publications, Birmingham, Alabama, 2002) is in effect a manifesto for the next wave of Islamist governments.

At one level, Qaradawi is a literalist in that he regards every word of the Koran as the word of God, which he sees as applicable for all times to come. But he also understands that an environment that supports critical thinking was one hallmark of Islam's golden age of scientific development (see 'Islamic era science'). Significantly, he has recently moved closer to philosopher Ramadan in his belief that Islamist governments should encourage self-criticism, that they should learn from failure, and that they have a duty to protect freedoms, including academic freedom and the freedom of any citizen to disagree with the state. "We want scientific thinking and the scientific spirit to guide our life in every way," he says. "It is against the scientific way of thinking to oversimplify complicated issues, or to view difficult problems with an alarming superficiality. Belief to us Muslims is not against reason or intellect."

Qaradawi is concerned that Islamist opposition movements are too literalist and are not doing enough to encourage independent thinking using reason, known in Arabic as *ijtihad*. "My worst fear for the Islamic movement is that it opposes free thinking for its followers and closes the door to *ijtihad*," he says. "If my fear turns into reality, then capable minds that can renew and innovate will escape from our

A long tradition

Seen in its historical context, the return of Islam to politics in countries with large Muslim populations is not surprising. It is a form of government that many experienced with few interruptions for nearly 1,400 years. Until the end of the First World War, most countries in the Middle East were provinces in the Ottoman Empire, and provincial governors applied variants of laws based on the Koran and on the life of the Prophet Muhammad.

The Ottoman Empire was the most recent incarnation

of the Islamic caliphate, whose rule began shortly after Muhammad's death in the seventh century, spreading Islam and Islamic government into Africa, Asia, the Middle East, and eastern and southern Europe.

Islamic law (sometimes called *sharia* law) is based on two principal sources: the Koran and a record of Muhammad's daily life



known as *hadith*. These are complemented by consensus among religious scholars, and independent reasoning by analogy. Among the Sunni and Shia schools of

Islam, the latter gives more weight to reasoning and critical thinking, which goes some way to explain the differences between Shia Iran and Sunni Saudi Arabia in bioethics, for example. **E.M.**

N. ISHTAYEH/AP/EMPHICS



Support for nuclear physicist Abdul Qadeer Khan, under house arrest for selling nuclear technology.

ranks, leaving behind those conservatives who can only imitate and who would like everything to stay as it is, regardless of how ancient it is." *Ijtihad* is sometimes called Islam's forgotten pillar. To others, it poses a threat to Islam by weakening its teachings.

Islamists have a reputation for looking inwards and shutting out the outside world, but they can look west when they need to, says Abdelwahab El Affendi of the University of Westminster's Centre for the Study of Democracy, in London, and chronicler of the rise of Islam in Sudanese politics. "Islamists that come to power on the back of 'we don't need the West' rhetoric end up becoming more pragmatic," he says.

Some Islamic thinkers are reaching out to the West in surprising ways. The prominent Turkish writer and columnist Mustafa Aykol has creationist views and publishes translations of US proponents of intelligent design. He has been building alliances with US faith-based groups such as the Discovery Institute in Seattle, Washington state. In an article for the US *National Review* last year he wrote: "Intelligent Design can be a bridge between these two civilizations. Muslims are discovering that they share a common cause with believers in the West."

In the late nineteenth century, Darwin's *On the Origin of Species* had a favourable reception in Muslim countries. But that is history, as books, pamphlets and films on creationism

are now more popular in Muslim countries, and pro-evolution scientists are afraid to speak out. Adults in Turkey, for example, are even less accepting of evolution than are those in the United States.

Nick Matzke of the National Center for Science Education, a not-for-profit organization based in Oakland, California, has debated intelligent design with Aykol in a Muslim online forum — a first for all concerned — but he thinks that Aykol's enthusiasm for the United States is unlikely to be reciprocated. American conservatives, he says, are not about to reconsider their views on Islam any time soon. "I find it peculiar that Muslims are adopting a doctrine from US groups that regularly bash Islam in a fairly vicious way," he says.

At Peshawar University, meanwhile, vice-chancellor Rashid is looking to increase direct links with foreign universities, having concluded an agreement to carry out teaching and research jointly with the University of Leicester, UK; the city of Leicester has a large British Asian population. Excellence in teaching, research and creative endeavour are the highest priority, Rashid says. But for him, Peshawar University's ultimate aim has to be a higher one. This is: "to love and serve the entire creation of the creator."

Ehsan Masood writes about science in developing countries.

See Editorial, page 1.

"It is against the scientific way of thinking to oversimplify. Belief to us Muslims is not against reason or intellect."

— Yusuf Al-Qaradawi

R. ANTROBUS/CORBIS

ART ARCHIVE/UNIV. LIB. ISTANBUL/DAGLI ORTI

1126–1198 Ibn Rushd (Averroës)

Spanish-born Islamic philosopher who tried to reconcile the contradictions between aristotelian ideas of studying nature through observation and reason, and religious truth. His writings and translations had considerable influence in Europe.



c.1259–c.1304 Maragha Observatory



One of the top three observatories in the Islamic world, this was built in Maragha in modern-day Iran. Astronomy was highly valued, partly for accurately predicting prayer-times and the Islamic lunar month. Maragha had a library of 400,000 books and a school of astronomy.

1201–1274 Nasir al-Din al-Tusi

Persian astronomer, mathematician and astrologer who worked at the Maragha Observatory. He introduced the 'Tusi couple', allowing Islamic scholars to greatly improve ptolomeic models of planetary motion.

1213–1288 Ibn al-Nafis

Damascus-born physician who worked in Cairo hospitals and produced the first recorded explanation of pulmonary circulation. But the mechanism remained largely unknown until William Harvey's work in the early 1600s.

1281–1923 OTTOMAN ERA

The Ottoman Empire spread from Anatolia into north Africa, Asia, the Middle East, and eastern and southern Europe.

1284 Al-Mansuri/Qalawun hospital, Cairo

Specialized institutions that treated disease for free and conducted research took root under Islamic rule, building on Roman efforts. The hospitals in Cairo and in Baghdad had wards for different illnesses. Clinicians took detailed case notes, which were collated into teaching manuals.



c.1304–1375 Ibn al-Shatir

Damascus-born astronomer and mathematician who developed new models of the Moon and planetary motion that eliminated problems with Greek models. Aspects of his work are identical to that produced by Copernicus.

1332–1406 Ibn Khaldun

Tunisian-born historian regarded as one of the first sociologists. His book, *Muqaddimah*, which is still in print, tries to explain why societies and economies change.

Ayla Arslan

THE DATA GAP

Statistics on scientific investment and performance are lacking across the Muslim world. **Declan Butler** analyses the best of what is available.

Stretching from Indonesia to Morocco, and from Uganda to Kazakhstan, countries with large Muslim populations are home to some 1.3 billion people. The Islamic world encompasses remarkable diversity in political systems, geography, history, language and culture (see page 20). But science in these nations is weak, with spending on research and development far lower than the global average. This much is acknowledged to be true, but what of the details behind the broad picture?

The official statistics database of the Organization of the Islamic Conference (OIC) reveals information on each of the 57 OIC nations on everything from arable land per tractor to Internet users, but you won't find any data on research¹. Science indicators for OIC countries are also scarce among data collected by the World Bank and United Nations agencies — largely a reflection of many of these countries' low level of interest in science.

To get a more detailed picture of how OIC countries measure up on science and technology, and of what patterns exist within OIC countries, *Nature* extracted science indicators



Can Islamic nations learn to capitalize on science and technology?

from official sources and reanalysed them for the OIC group as a whole, creating an overall picture of science and technology indicators for OIC countries.

Recognized authorities such as UNESCO or the World Bank Development Indicators have few reliable data on science spending in most OIC countries. But combining these data for 20 OIC nations covering 1996–2003 gives the average annual spend on R&D as 0.34% of GDP, much lower than the global average over the same period of 2.36% (refs 2,3).

Although many OIC countries are among the world's poorest, with almost half being developing countries, their spending is consistently less compared with the national average across a range of income brackets. The exceptions are Malaysia and Turkey, whose spending is comparable to other moderately wealthy nations. Nowhere is the OIC deficit greater than in the oil-rich nations; Saudi Arabia and Kuwait, for example, spend less proportionately on research than the poorest OIC countries (see chart below, and page 28).

Part of the explanation lies in spending priorities. Many OIC countries, particularly the richest, spend more on armaments than on science, education or health³. Six of the world's top ten military spenders as a share of public spending are OIC countries: Kuwait, Jordan, Saudi Arabia, Yemen, Syria and Oman (each spending above 7% of GDP on arms in 2003). African OIC countries, in contrast, tend to spend proportionately less on the military.

A tough lesson

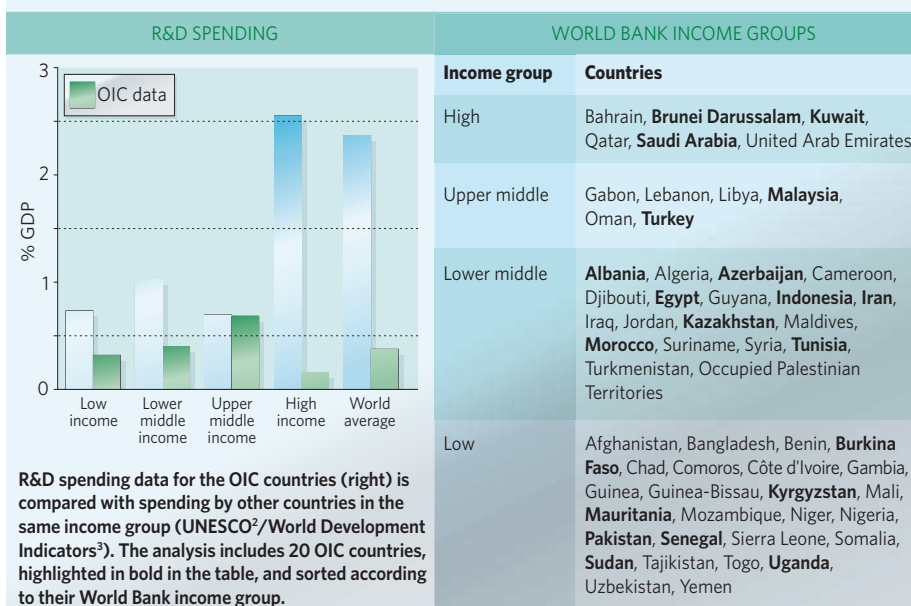
Although the science budgets of the OIC countries are all near the bottom of the world league, their spending on education is more variable. Malaysia, Saudi Arabia and Yemen's relative education budgets are among the world's highest. Morocco, Tunisia and Iran also spend respectable sums on education. All six were among the world's top 25 spenders on education in 2002 (ref. 3).

The OIC countries' performance in education also varies more widely, according to the World Bank's 'education index', suggesting that many have the human resources that could exploit greater investment in science and technology³.

But of the 20 poorest performers on this score, 15 are OIC countries, including many African nations, Bangladesh and Pakistan. The OIC countries' low investment in science and technology is also reflected in a poor scientific output, including low levels of scientific articles and numbers of researchers.

World Bank Development Indicators for 1996–2003 record numbers of researchers per million people for 19 OIC countries³. These OIC nations cluster at the bottom end of the global scale. The top global performers (Finland,

R&D SPENDING BY THE OIC COUNTRIES RELATIVE TO INCOME GROUP



Iceland, Sweden and Japan) all have above 5,000 researchers per million people. The highest scoring OIC country is Jordan, with 1,927 researchers per million people; the OIC average is 500.

Of the 28 lowest producers of scientific articles, as recorded by the US National Science Foundation⁴, half are OIC countries. In 2003, the world average for production of articles per million inhabitants was 137, whereas none of the 47 OIC countries for which there were data achieved production above 107 per million inhabitants⁴. The OIC average was just 13.

Moreover, over the past two decades the number of papers produced by 24 OIC nations has remained flat or declined, albeit with some striking exceptions⁴ (see chart, right). Turkey's publication rate per year has grown from around 500 in 1988 to more than 6,000 in 2003. The other rising star is Iran, which from a low base of less than 100 articles per year a decade ago now produces nearly 2,000. Both countries have eclipsed Egypt, previously the most prolific of all OIC states in scientific publishing, which grew only slowly between 1988 and 2003.

Scientific divide

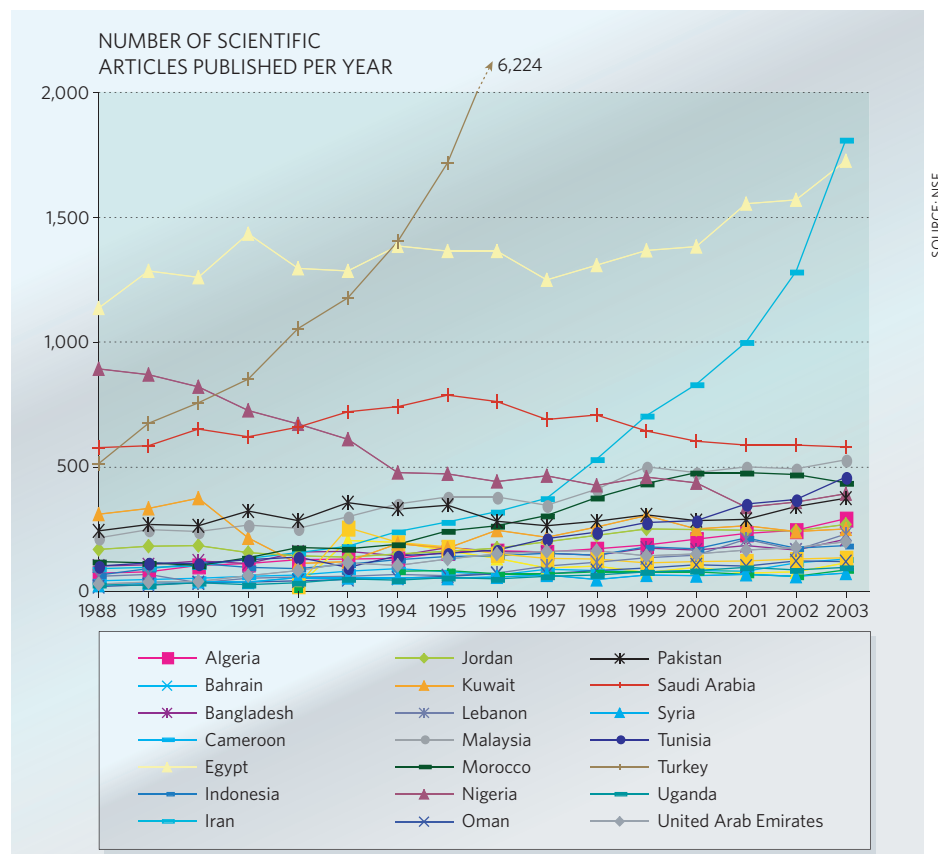
The articles published by OIC countries show striking disciplinary and geographical differences³. Health and social sciences get little attention except in south Asia and African countries. Chemistry and physics papers make up greater shares of total output in central Asian countries, life sciences in North Africa, Indonesia and Malaysia, as well as Saudi Arabia, and engineering predominates in most Middle Eastern and north African countries.

The OIC countries produce so few patents that they are invisible on a bar chart of comparison with other countries³. This lack of technological competitiveness translates into low rankings in terms of high-tech exports as a percentage of total exports — with one exception, Malaysia, which ranks fifth worldwide with 58% high-tech exports, alongside Singapore and the Philippines. Otherwise, Indonesia (14%) and Morocco (11%) are the only OIC countries with high-tech exports greater than 10%.

Given such diversity, it is hard to identify trends across the Islamic world as a whole. But there are signs of hope when looking at the OIC's top and bottom performers.

Turkey, for example, is not rich in oil, but is the most scientifically successful of the Muslim states. Turkish intellectuals attribute this to the 1923 revolution, which led to the creation of a constitutionally secular state.

For decades, Turkey has aimed for membership of the European Union (EU); formal negotiations began last year. Negotiations require even closer structural alignment with EU



member states, and since 2003 science funding has more than trebled.

Modern Turkey grew from the ruins of the Ottoman empire. Mustafa Kemal Atatürk, Turkey's founder, was keen for his country to catch up with the West. Universities switched from using Arabic to the Roman alphabet, ensuring easier access to Western writings, and Atatürk initiated a nationwide campaign to raise literacy rates.

Balancing act

The Kemalist legacy has fostered Western-style scientific organizations and policies. But the insistence on removing religion from public life introduced restrictions of its own. For example, women are banned from wearing headscarves in government-funded universities — something that may have to change if Turkey joins the EU. Yet Turkey has respectable participation by women in academic life — higher than in some EU countries.

As for the bottom performers, sub-Saharan Africa accounts for 21 of the 57 OIC countries and, with the exception of Cameroon and Gabon, are among the poorest.

The ongoing economic recovery of sub-Saharan Africa is "one of the most remarkable stories of the past five years", according to the 2006 World Bank Development report³. The latest available data, for 2004, show five years of growth, after two decades of decline. Most was down to oil, but agriculture was also important. Among the OIC countries, Chad's economy grew more than 10% per year and Nigeria's 6%.

But overall economic levels remain poor,

with growth rates usually far below the 7% minimum considered necessary to begin meeting the Millennium Development Goals. Several countries, including Côte d'Ivoire and Gabon, still show negative growth.

Not surprisingly, science is correspondingly weak, and although hard data are few, Islamic countries in sub-Saharan Africa have about 20 researchers and engineers per million people, compared with about 250 in Latin America. One potential bright spot is Nigeria, which announced in July that it is considering ploughing its oil revenues into a US\$5-billion endowment fund for science and technology, which would make it Africa's biggest science spender by far⁵.

If others follow where Turkey and Nigeria lead, these science indicators might look very different in ten years' time. As a first step, more OIC governments need to gather data on science, producing figures that are sufficiently reliable to appear in international statistics databases.

Declan Butler is a reporter for Nature based in Paris. Additional reporting by Ayla Arslan.

1. The Statistical, Economic and Social Research and Training Centre for Islamic Countries (SESRTCIC) www.sesrtic.org/statistics/bycountry.php
2. UNESCO Statistics Division <http://stats.uis.unesco.org/ReportFolders/reportfolders.aspx>
3. World Development Indicators (WDI), 2006 <http://devdata.worldbank.org/wdi2006/contents/index2.htm>
4. National Science Board. Science and Engineering Indicators 2006. (National Science Foundation, Arlington, VA) www.nsf.gov/statistics/seind06/
5. Giles, J. *Nature* **442**, 334 (2006).

See Editorial, page 1; Commentaries, pages 33, 35.

OIL RICH, SCIENCE POOR

The wealthy Arab states offer scant support for science and technology.

Jim Giles finds out whether this indifference to research is likely to change.

When *Nature* surveyed the prospects for science in the Arab world in 2002, our reporter picked out three subjects in which the region excelled¹. One was, and still is, important: desalination technologies to combat water shortages. But the other two highlight the region's threadbare research record. Camel reproduction and falconry research might excite Arab sports enthusiasts, but they are unlikely to set the scientific world on fire.

The monarchies of the Gulf are the richest of all Muslim nations, but little of that wealth is spent on research. Saudi Arabia, Qatar and Kuwait spend about 0.2% of their gross domestic product (GDP) on science — less than one-tenth of the developed-country average of 2.3% and about a third of that spent by less wealthy Iran. The oil monarchs have the financial clout to launch major research efforts, but have yet to do so.

"The very rich countries are less concerned because they are sitting pretty on oil reserves," says Nader Fergany, director of the Almishkat Centre for Research in Cairo, an independent social-sciences research organization. "The nature of wealth from natural resources is that it does not require a great level of ingenuity," Fergany notes that even in directly relevant science such as petroleum technology, most innovation happens outside the Gulf.

Oil futures

But some Gulf leaders do see investment in science and technology as a way of creating an economic future when their oil reserves dry up. Among scientists trying to invigorate science in the Gulf, there is a sense that change is possible. "We are now at an inflexion point," says Samir Hamrouni, director of research and development at the Arab Science and Technology Foundation in Sarjah in the United Arab Emirates. "Science is being seen as an alternative to natural resources."

The origins of the current underspend are easy to see. The European colonial powers that ruled much of the Gulf until the middle of the twentieth century invested almost nothing in indigenous higher education or research. Oil revenues transformed the region, but the money kept flowing without the need for major investment in education and science.



K. JEBREIL/AP

Easy option: the Gulf oil states imported most of the know-how they needed to keep the oil flowing.

The latest statistics collected by COMSTECH, the science and technology committee of the Organization of the Islamic Conference, show little change². The annual output of scientific papers from Saudi Arabia, which generates almost as many papers as the other monarchies combined, was static between 2000 and 2005. Even in desalination technology, investment has been limited. The Middle East Desalination Research Center in Muscat, Oman, set up in 1996 to encourage research cooperation in the region, is currently limping along with a budget of just US\$2 million a year.

The next five years might see more change. In Qatar, the country's head of state, Emir Hamad bin Khalifa Al-Thani, has created an endowment that generates millions of dollars in research funding every year. He has also imported Western science policies, such as competitive grant systems based on external peer review, and is forming partnerships with universities in the United States and Europe³. Environmental science, computing and biomedicine are priorities.

If Qatar's new research centres attract scientists and students, they might prompt its neighbours into action. "Once it develops, other countries will start to think about it," predicts Mohamed Hassan, executive director

of the Academy of Sciences for the Developing World (TWAS), based in Trieste, Italy.

Of those neighbours, Saudi Arabia is making a slow start, having approved a new national science and technology development plan in 2002. Its priorities are defence, and oil and gas technology, but there is also a commitment to devote 1.6% of the nation's GDP to R&D by 2020. Both Saudi Arabia and Kuwait are each investing around \$2 billion in higher-education institutes that include research centres.

Such initiatives generate excitement — and some scepticism. Fergany questions whether the oil monarchies are willing to make the economic and structural changes needed to translate research into innovation. It is also unclear whether the oil-state rulers want to foster the atmosphere of critical enquiry that science needs. Only in the long run, say advocates of reform, will it become clear whether the current commitment is genuine. "Is it just for the moment, or is it really important?" asks Hamrouni. "It depends on our politicians." ■

Jim Giles is a reporter for *Nature* based in London.

1. Masood, E. *Nature* **416**, 120–122 (2002).
2. *Status of Scientific Research in OIC Member States* (eds Naim, S. T. K. & Atta-ur-Rahman) (COMSTECH, 2005); online at www.comstech.org/htm/policy.htm.
3. Giles, J. *Nature* **441**, 132 (2006).

See also Commentaries pages 33 and 35.

Q&A

The reformer

Mostafa Moin is a paediatrician and medical researcher who has served as Iran's minister for higher education and for science. He was a reformist candidate in Iran's presidential election last year, which was won by religious conservative Mahmoud Ahmadinejad. **Declan Butler** asks Moin about the prospects for science in Iran.



R. HOMA/AND/REUTERS

How do you see the interplay between science, religion and reform in Iran?

Iranian society has long been deeply religious, even before the Islamic era began in 651. But over the past 150 years, Iran has also pioneered struggles for freedom and opposition of dictatorship.

In the past few decades, reformers and religious neo-intellectuals with a common attachment to the principles of a civil and democratic society have cultivated democratic structures in Iranian society.

These will undoubtedly prepare the ground for greater scientific development and help a knowledge-based society to materialize. Islam itself is not anti-science, but I have always been concerned that superficial, narrow-minded and non-democratic interpretations of Islam — and the political behaviour of certain traditional administrators — risk having a negative impact on both scientific development and social reforms in Iran.

What is your assessment of Mahmoud Ahmadinejad's track record on science, academic freedom and social reforms?

The new government has overlooked the science and higher-education sectors in the 14 months it has been in power. It has also replaced almost all the university chancellors, and senior research and higher-education officials. When I was minister of science, research and technology, senior university officials were elected by the academic staff; the new government appoints them directly.

There is no doubt that this political control has resulted in purging, restrictions and criticism of independent forces. Renowned academics have been forced into retirement, and repression of politically active students and student organizations is escalating. International scientific exchange has not been immune from these narrow-minded

approaches: students are no longer sent abroad and sabbaticals are restricted.

The government has cracked down on reformist newspapers, activists and political parties. The *Shargh Daily*, for example, one of the highest-circulation reformist newspapers, was shut down in September. I've also heard that admission of female students to universities is being restricted; I hope my information is incorrect.

How optimistic are you that Iran can find a route to political reform?

In the short term, it can't be predicted whether things will get better or worse. But I'm optimistic for the medium term. There is a global movement towards greater civil rights, and an expansion of democracy. It is this, and the growing awareness of the Iranian people of these issues, the vigilance of its youth and women, that will shape Iran's future.

What are the biggest obstacles to improving Iran's international isolation?

International scientific cooperation with Iranian universities and scientists has greatly increased in the past few years. But it's not surprising that the nuclear crisis and the unscientific and unsustainable policies of the new government may have overshadowed this process. Current foreign policy is based on confrontation, and is shifting away from former president Mohammad Khatami's 'dialogue of civilizations'. This is the main obstacle at present.

What should colleagues elsewhere think about Iran's nuclear programme?

I am sure that my learned academic colleagues abroad would not accept discrimination against Iran's legitimate right to the peaceful use of nuclear energy within the framework of global regulations. The problem is the

accusations by the ruling neo-conservative radicals in the United States over programmes of weapons of mass destruction in Iran, while proposing double standards within the Middle East and the rest of the world.

The same accusations have been made before, but experts, scholars and international institutions proved them unfounded. If the United States were to implement unilateralist, violent and non-negotiable policies against Iran, this would be a great catastrophe, increasing regional and global crises, expanding terrorism and consolidating dictatorships.

What were the major defining events in Iranian science over the past two decades?

The expansion of higher education in the country, with university students increasing from 400 to more than 3,000 students per 100,000 people, between 1979 and 2000. The student-lecturer ratio has also improved from about 36:1 in 1989 to 18:1 in 2006.

I regret that the structural reform in higher education in Iran has remained unfinished, and that the autonomy of universities and academic freedom were not institutionalized.

What are your own plans for the future?

In the past two to three years I have founded two non-governmental organizations, the Association for Scientific Development of Iran and the Iranian Association for Ethics in Science and Technology. As president of the Immunology, Asthma and Allergy Institute at Tehran University of Medical Sciences, I'm active in teaching and research.

I have respected the commitment I made to the Iranian people during last year's election, with the creation of the Democracy and Human Rights Front. I want more than ever to strengthen civil and scientific institutions and structures, particularly for the young. ■

Dishing a modern myth about microbes

SIR — We are indebted to Lynn J. Rothschild for debunking, in *Journal Club*, some microbial myths (*Nature* **443**, 249; 2006). Her prime concern is that many microbiologists have simply given up trying to culture many of the more recalcitrant microbes living in the natural environment, in favour of resorting to DNA-sequencing methods.

There is a lingering and widespread presumption that DNA-sequence information will make a huge difference to our knowledge of microbes in the natural world, but we, like Rothschild, beg to differ. She argues for a greater effort to learn how to culture the 'unculturable' microbes. We would go further, and suggest that, having done so, we should devote greater effort to describing species by their phenotypic properties. Characterization of the phenotype, rather than the genetic sequence, is the key to gaining deeper and better understanding of microbial diversity in the natural world.

Tom Fenchel*, **Genoveva F. Esteban†**,
Bland J. Finlay†

*Marine Biological Laboratory, University of Copenhagen, DK-3000 Helsingør, Denmark

†CEH-Dorset, Winfrith Technology Centre, Dorchester, Dorset DT2 8ZD, UK

The daunting process of MIAME

SIR — Michael Eisenstein, in his Technology Feature "Quality control" (*Nature* **442**, 1067–1070; 2006), reports recent improvements in microarray technologies. He discusses the establishment of systems for "clearly defining how [microarray] data were obtained", such as the MIAME (minimum information about a microarray experiment) annotation standards.

There are two problems inherent to the use of such standards. First, microarrays generate datasets whose intrinsic dimensionalities are several orders of magnitude greater than the information content of the standards. Second, one cannot know *a priori* all the biological factors that might affect the results reported by microarrays.

In terms of the first point, the prospect of defining 30,000-odd different MIAME descriptors — the routine size of microarrays — would be daunting under any circumstances. The second point is neatly illustrated by work described in the same issue (J. Molinier, G. Ries, C. Zipfel and B. Hohn *Nature* **442**, 1046–1049; 2006). This study reported the unexpected finding of transgenerational, non-genetic memory of stress events in the flowering plant *Arabidopsis thaliana*. Although transcriptome analyses in this case did not

detect differences in global gene expression (possibly due to lack of sensitivity), the question remains as to whether one should be required to record the prior generational treatments of plant lines in microarray experiments — or in any experiment. Such a requirement appears equally daunting.

How should we proceed? Reducing the costs of microarray technology so that experiments can be readily reproduced across laboratories seems a reasonable approach. Relying on minimal standards of annotation such as MIAME seems unreasonable, and should be abandoned.

David W. Galbraith

Bio5 Institute and Department of Plant Sciences, University of Arizona, Tucson, Arizona 85721, USA

No room for complacency on drug resistance in Africa

SIR — Your News Feature "Staying the course" (*Nature* **442**, 617–619; 2006) offered a welcome dose of optimism in the debate about mass deployment of antiretroviral therapies in Africa, by arguing that this had not yet created an epidemic of drug-resistant HIV. But although everyone in the field has been delighted by initial studies suggesting high adherence, we must not be complacent.

The News Feature cited studies showing that 70–80% of patients on antiretroviral drugs had undetectable viral loads; the converse is that 20–30% have detectable viral loads. The spread of resistance is determined not only by compliance — which, as you point out, has been good so far in most settings — but also by factors more specific to Africa, such as interactions between HIV and TB drugs, or by limitations of drug-supply infrastructure. The predominant use of clinical monitoring in African programmes may allow mutations to accumulate (DART Virology Group *AIDS* **20**, 1391–1399; 2006), potentially making resistant strains more transmissible. If virological failure accompanies the roll-out of these drugs, it could rapidly increase with mass deployment: see how fast this might happen at http://pcwww.liv.ac.uk/hastings/HIV_models/2006_HIVdynamics.xls. Scientists must pool data from roll-out cohorts (functioning as sentinel sites), not only to assess resistance but also to gather data on adherence, pharmacokinetics, drug interactions and other factors that determine a drug regimen's useful lifespan.

The consequences of widespread use of nevirapine to prevent mother-to-child HIV transmission are not yet clear, and a public health fall-back plan is urgently needed in case large-scale resistance emerges. Well-funded education programmes are vital to help people minimize their risk of infection, encourage them to undergo the counselling and testing they need to gain access to

antiretrovirals, and overcome the stigma of AIDS. The success of initial antiretroviral deployment is encouraging, but we should not underestimate the challenges that remain.
Ian M. Hastings*, **David G. Lalloo***,
Saye H. Khoo†

*Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK

†Department of Pharmacology and Therapeutics, University of Liverpool, Liverpool L69 3BX, UK

Brief: goodbye to a quirky perspective on science

SIR — I am writing to protest at your decision to cancel the Brief Communications section (*Nature* **443**, 246; 2006). I have always enjoyed the high-quality yet rather quirky things published there. Where else would I learn about synchronized hand-clapping, the photonic band structure of beetle exoskeletons or mathematical models of bilingual societies? I turn to the Brief Communications section to see ways in which rigorous scientific approaches can shed light on very novel questions.

I have not yet come across something appropriate in my own research to submit as a Brief Communication, so I suppose that, having done nothing to alleviate the quality problem, I have no right to complain. (I am working in my spare time on a result that fits the spirit of the section, being a mathematical proof that there's a 'best' way to vote, but I can't think of a way to reduce it to a single page.) Still, I doubt I'm the only reader who's disappointed by this loss.

Brief Communications have always served to remind me that science can be both excellent and quirky. When I was teaching optics and discussing novel optical phenomena in the animal kingdom, I would show students the article "Opal analogue discovered in a weevil" (*Nature* **426**, 786–787; 2003). The second page included a Brief Communication titled "Health benefits of eating chocolate?". Without Brief Communications, how would I be able to teach students both that insects are talented optical engineers and that chocolate may be good for your health?

Alex Small

Laboratory of Integrative and Medical Biophysics, National Institute of Child Health and Human Development, Bethesda, Maryland 20892, USA

Brief Communications (RIP) and the soul of wit

SIR — "What a pity," he said — briefly.

Jeremy Wolfe

Visual Attention Lab, Brigham & Women's Hospital, Cambridge, Massachusetts 02139, USA

COMMENTARY

Steps towards reform

Building a knowledge-based society in today's Arab world depends on overcoming primarily political obstacles to progress. **Nader Fergany** analyses the reforms required for an Arab renaissance.

Building a knowledge-based society in the Arab countries would mean reclaiming one of the epic achievements of Arab history. At the zenith of Arab-Muslim civilization, from the eighth to the thirteenth century, the Arab world succeeded in building a knowledge culture in which Arabic was the language of science, and knowledge production flourished.

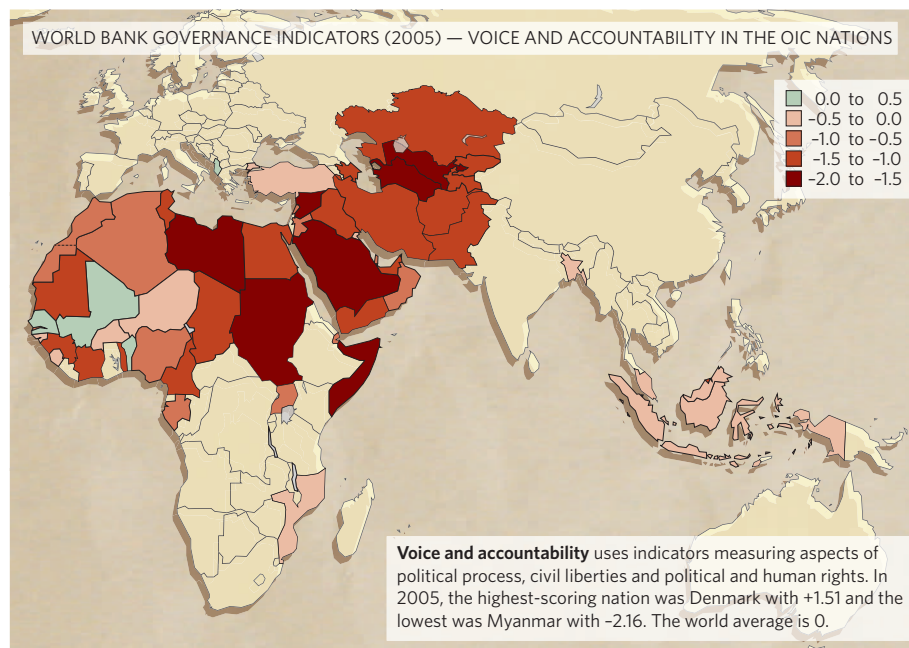
But building a knowledge society in today's Arab world is a significant challenge. Arab countries spend vastly less on science than most other nations. Investment in education is also lagging, with students spending far fewer years at school than their counterparts in the east Asian tigers, for example. The Arab region also has far fewer researchers and engineers, and produces fewer scientific publications, than any other world region, apart from sub-Saharan Africa.

Deficiencies in science and education investment are secondary obstacles to progress — the primary impediments are political. That is the conclusion of a series of Arab Human Development Reports (AHDR) published between 2002 and 2005 (<http://cfapp2.undp.org/rbas/ahdr.cfm#>). The fourth and final report in the series, devoted to the 'Rise of women in the Arab world', is scheduled for launch in December. The reports, of which I am the lead author, are written by Arabs for Arabs, by a team of more than 100 scholars and experts, sponsored by (but independent of) the United Nations Development Programme.

Knowledge society

The series formulates a strategic vision aimed at restructuring the region from within to achieve an Arab renaissance built around a knowledge-based society. In this vision, knowledge itself is crucial to the functioning of an open society, and is intimately linked with human development. Establishing institutions of good governance is critical to the expansion of human freedoms, justice and dignity.

The first AHDR report, published in 2002, identified three major barriers impeding human development in the Arab region: deficits in the acquisition and production of knowledge, in the empowerment of women, and in freedom. These deficits exist on three interlinked levels, from the rights of the individual to those of subgroups, and, equally important, national liberation and self-deter-



mination. Moreover, the report concluded that all three barriers are inextricably linked — greater freedoms are a prerequisite for creating an environment conducive not only to the creation of knowledge, but also the starting conditions for a profound reform of governance ensuring justice and human dignity.

Subsequent AHDR reports have endeavoured to investigate these three topics in depth. 'Building a Knowledge Society', published in 2003, defines societal knowledge as including not just the natural and exact sciences, but also the humanities and social sciences, and artistic and literary creativity. This broad definition of knowledge is important, as it underpins a vision of a society in which the production and dissemination of knowledge become central to its every aspect.

A strategic vision for creating a knowledge society in Arab countries may, at first, seem fairly free of political implications, dealing as it does with the relatively neutral issues of education and science. But it is in fact a deeply political issue. It is no coincidence, for example, that a central pillar of the reports' vision is the "total respect for the key freedoms of opinion, expression and association, and guaranteeing these through good governance" (see 'Pillars of knowledge', overleaf).

As World Bank governance data for 2005 reveal, most of the nations belonging to the Organization of the Islamic Conference perform badly on indicators for voice and accountability (measuring political, civil and human rights, above) and for effective governance (measuring the quality of public-service delivery and bureaucracy competence, overleaf).

Good governance

Political reform is required therefore at national, regional and global levels. At the national level, Arab regimes are rarely representative of, nor accountable to, the people. Power, in the form of political authority and wealth, is concentrated in the hands of a few. Most of the public are marginalized and live in poverty — the antithesis of freedom.

At the regional level, current efforts at cooperation between Arab countries have failed to capitalize on the immense potential offered by closer Arab integration, despite these countries having the advantages of a common culture, history and language. This failure is all the more disheartening given that Arab countries face serious regional and global challenges that could be better addressed united. The potential for greater Arab cooperation has, for example, largely gone untapped in key areas

of opportunity such as education, mass media and in research and development in fields of particular relevance to the Arab region.

Globally, knowledge has been transformed from the 'public' good it was for much of history to a 'private' good governed by the profit motives of multinational corporations. Historically, the fruits of Arab-Muslim science were freely given to the West, enabling the European Renaissance. In this light, any attempt to restrict less-developed countries' access to knowledge, for example through international intellectual-property rights agreements, should be seen as a historical reversal of one of humanity's noblest traditions: the free sharing of knowledge.

Human development and freedom in Arab countries are also being restricted by the failure of global governance to help resolve fairly the intensifying conflicts that beset the region, foremost among which is the foreign occupation of Palestine and Iraq.

The Arab world has substantial potential human capital. Its scientists innovate as best they can, despite an environment that is hostile to investment and has weak institutional support for research — both of which favour an ongoing brain drain. A first step would be to draw on the potential of the Arab scientific diaspora to contribute to the creation of a knowledge society. There need to be policies that offer incentives for Arab scientists working abroad to return to their home countries.

Cultural renaissance

But the Arab development crisis is so acute, complex and multifaceted that a knowledge renaissance cannot be constructed in isolation, and will be impossible in the absence of broader reforms touching almost all aspects of Arab society — in cultural constructs, social and economic systems, and above all in the political structures at the national, regional and global scales.

In other words, partial reform is no longer enough. The developed world is rapidly forging ahead in creating knowledge-intensive societies. If the Arab world does not quickly reform, the asymmetry of world knowledge production will continue, and the Arab countries will be forever marginalized.

At the same time, a knowledge renaissance demands that Arabs open up more to the world and other cultures. Arabs have contributed greatly to human culture, and must now regain the honour of effectively participating in it. Isolationism can only lead to stagnation and impotence.

World order is also indisputably in need of reform to reduce inequalities, injustice and oppression. By reinforcing Arab cooperation, and actively engaging other cultures, Arabs can help accelerate a global enlightenment.

A radical reform of the power structure in Arab countries is needed. Several scenarios can be contemplated. The first is the 'impending disaster' scenario. Here the status quo in Arab countries is maintained by corrupt governance in regimes clinging on to power. The most likely outcome will be violent social conflict. This may result in a change in power structure but probably at an unacceptably high price.

Similarly, reforms imposed from outside the Arab world cannot achieve acceptable social change, as ultimately they will serve the interests of those imposing them, and not the interests of the Arab people. This inevitably results in legitimate resistance.

Peaceful reform?

The only alternative is reform from within Arab countries themselves. This would require the emergence of a vigorous civil society capable of leading a peaceful process of political struggle. But Arab society remains stuck because of restrictions on freedom and the prevailing climate of corruption.

Pillars of knowledge

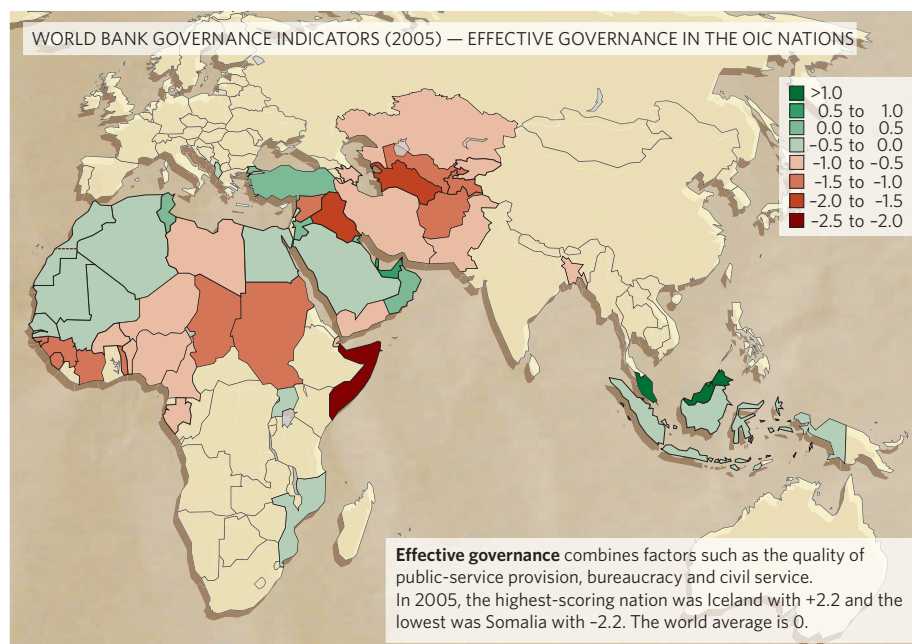
The strategic vision laid out in the Arab Human Development Report for creating a knowledge-based society in Arab countries is organized around five main pillars:

- Total respect for the key freedoms of opinion, expression and association, and guaranteeing these through good governance.
- Ensuring high-quality education for all, with special attention given to early childhood education and higher education, and to life-long continuing education.
- Introduction of research and technological development in all societal activities, and an accelerated transition to the information age.
- Rapid transition to a society in which knowledge becomes the source of value and the organizing principle of human activity.
- Establishing a general Arab knowledge model that is authentic, open, enlightened and based on five factors: a return to true religion; promotion of the Arabic language; furthering of historic Arab knowledge; celebration of cultural diversity within nations; and opening up to other societies and cultures.

Will existing regimes ever allow greater freedom of assembly and organization in civil and political society? Small Muslim countries, such as Albania and Senegal, have shown over the past decade that it is possible to convert a negative indicator for freedom of expression into a positive one. Is peaceful reform possible in the larger populous Muslim or Arab nations? The result would be a historic renegotiation of the existing structure and exercise of power.

Among the 22 Arab League nations only three — Algeria, Bahrain and Iraq — have made any positive improvement in voice and accountability over the past decade, and this from historic lows. Many more Arab nations have seen negative trends and a worsening of freedom of expression that makes the 'impending disaster' scenario much more likely.

Hope is pinned on a reform process that begins with total respect for the key freedoms of expression and association. This would allow the emergence of a vigorous civil society capable of leading the required transformation. I believe that the seeds are there; in mushrooming civil movements protesting against despotism and corruption, and clamouring for deep-rooted reform. In Egypt, for example, the past few years have witnessed the emergence of several increasingly effective protest movements — such as the broad 'kefaya' movement — that have inspired similar movements in other Arab countries, such as Yemen and Libya. If we fail to cultivate them it will be an irrevocable loss of a historic opportunity. ■ Nader Fergany is director of the Almishkat Centre for Research, Giza, Egypt.



See Editorial, page 1; News Features, pages 19–29.

Where are the new patrons of science?

Muslim nations must take a big leap forward in developing science and technology to catch up with the rest of the world, argues **Herwig Schopper**, or they risk falling behind in the global economy.

Spending on research and development in the Islamic world is an order of magnitude below the global average. According to a series of highly self-critical reports assembled by Arab scholars, and released by the United Nations¹, countries in the Arab world are finding it hard to improve the situation. As commentators from that region have noted^{2,3}, they continue to fall behind not only the developed countries in the West but also emerging nations in east Asia. In the past decade, Taiwan and South Korea have shown a breathtaking expansion of science and technology coupled with rapid economic growth. Can a similar leap forward occur in the Muslim world?

Lack of funding is a major obstacle but not the only reason for the appalling status of science in the wider Muslim world. Certainly for poorer nations, lack of resources is a problem, and they may view science as a luxury they cannot afford. Governments mostly exacerbate the problems of low investment because they lack the strategies to foster science and to prioritize research projects.

One of the biggest challenges facing Muslim countries is that science is too often viewed as a commodity that can be separated from the thought processes that led to it. Believing that oil money can simply buy Western technology, wealthy Arab states do little beyond consuming science and technology products. But to flourish, science and technology need a cultural base that can only be acquired by science education and by undertaking research programmes. This effort requires, above all, a return to the patronage of the past, when science had support at the highest echelons of Muslim society.

Another factor is poor integration within the international scientific community. In 2000, under the auspices of UNESCO (the United Nations Educational, Scientific and Cultural Organization), an international laboratory for synchrotron radiation for the Middle East and the Mediterranean regions, SESAME, was founded to promote science and technology in the region, to improve international cooperation and to bring nations together in peace. The laboratory is in Jordan, close to Amman. SESAME was created according to the model of CERN, the European Laboratory for Particle Physics near Geneva. Because of my experience with CERN,

"Today's Muslim societies have generated few scientists of international repute."



Falling on deaf ears? The president of Sudan, among others, has called for greater support for science.

I was asked to chair the SESAME council, the governing body of the organization. That is how I got a chance to understand at first hand the problems of the Middle East, and the different mentalities and ways of operating.

After many discussions with people from the region — scientists, administrators and politicians — my main message would be that unless drastic changes occur, it will fall further behind the rest of the world, despite its great culture, its human capabilities and its relative wealth. Many reasons have been given for the unsatisfactory situation in science and technology at present — including conflicts and economic sanctions — but no excuses will help. The competition in other parts of the world is progressing fast. A big jump in development is needed.

Muslim science had a golden age between the eighth and thirteenth centuries (see 'The golden age', overleaf, and timeline, page 23) and then declined. There are external reasons for the decline of Muslim science, including the Mongol invasion in the thirteenth century, but internal factors, such as growing isolation, growth of authoritative regimes, discouragement of innovation and restrictions on freedom of expression, were probably more decisive.

In general, there was a shift, starting around 1100, from a rational and tolerant attitude to a more conservative school of thinking that denounced philosophy and rationalism.

The older developed countries of Europe took some 150 years to evolve from an agriculture-based society into an industrialized one. But there is not enough time to repeat the European experiment in the Islamic world. If Muslim countries were to follow a similarly slow path to modernization, they would be doomed to insignificance in the global economy. There has to be a rapid leap forward to catch up with the industrialized world. That this is possible has been shown by countries such as Taiwan and South Korea, and more recently by China.

How to leap forward?

Above all, the mentality of political leaders must change to show more of a commitment to science. Political leaders in many Islamic nations simply fail to appreciate the importance of scientific research to their countries' development. This is not to ignore the real poverty of many Islamic nations, or the positive steps taken by some countries, such as Jordan, Pakistan, Iran and Turkey, which may still inspire others to action. And there are pronounced differences between the 57 Islamic countries that make up the Organization of the

N. NASSER/AP

Islamic Conference (OIC) that are discussed in more detail elsewhere in this issue. In short, not all of the criticisms or recommendations expressed here can be applied to every Muslim nation. Nevertheless, all Islamic countries can do more to support science and technology. Socio-economic development cannot depend solely on natural resources. Nowadays, knowledge has become a major driving force of world economics.

Many proposals have been made by individual Muslim leaders or organizations to support science and technology. Unfortunately, most of these have had little effect. World Bank Development Indicators for 1996 to 2003 reveal that Muslim countries on average spend less than 0.4% of their gross national product on research⁴. This figure compares with a world average of 2.36%. It is difficult to convince sponsors from other parts of the world to help the Muslim world if their own efforts are not stronger.

Sometimes the initiatives are full of good intentions, but the results are too often disappointing. In May 2002, the OIC's Standing Committee on Scientific and Technological Cooperation (COMSTECH) proposed that the Islamic Development Bank come forward with at least US\$1 million annually to upgrade some selected research institutes. At an Arab League summit of 22 nations in Khartoum, Sudan, in March, members agreed to collaborate more closely to increase science funding and encourage public-private research partnerships. Omar Hassan al-Bashir, president of Sudan, opened the summit with a call to put science at the heart of nations' strategic plans. He suggested that the increasing revenues from oil production should be used to fund science and technology development. To date, none of these statements has led to the concrete actions hoped for.

In the meantime, the Arab world is suffering an emigration of intellectuals to the West. According to a 2004 report by the Gulf Centre for Strategic Studies in Cairo, each year the Arab countries lose 50% of their newly qualified physicians, 23% of engineers and 15% of scientists, mainly to Britain, the United States and Canada⁵. In addition, some 45% of Arab students studying abroad do not go back to their countries after graduating.

Today's Muslim societies have generated few scientists of international repute, despite accounting for roughly one-fifth of the globe's population. Only two scientists from Islamic states have won Nobel prizes: Abdus Salam, a Pakistani (for physics in 1979) and Ahmed Zewail, an Egyptian (for chemistry in 1999). Both carried out their research outside Islamic countries.

Scientific progress in academia in Muslim countries is often hampered by isolation and a relatively immature university system.



Hope for the future: SESAME under construction in Jordan.

Promotions in institutions are often based on trustworthiness rather than merit. Salaries are so low in some countries that people have to find additional income. An excessively bureaucratic system often stifles innovation.

A few recommendations

Although my knowledge of the situation in Muslim nations is based on my personal experience in the region, I would not dare to make such negative statements on my own were they not confirmed by Muslim scientists. When I have presented my thoughts to people from these countries I have found support for many of the recommendations offered below.

First, the importance of research in contributing to the overall welfare of Muslim societies needs to be recognized, which in the short term will require a return to patronage at a high political level. This is possible, as some examples prove — the strong support of science and education by King Abdullah II in Jordan and by President Pervez Musharraf in Pakistan. Political leaders need to provide sufficient resources

for basic scientific research, which should be considered as investments rather than expenditures.

Inward investment is crucial because the infrastructure for research within individual countries is meagre and legal frameworks for innovation are largely non-existent. Building scientific infrastructure from the ground up is important for establishing procedures of competitive evaluation and for learning how to set priorities. National networks for research, including better electronic communication systems,

can greatly improve cooperation between scientists.

Second, Muslim nations need to strengthen international links, in addition to purely national or bilateral cooperative projects. Excellence in scientific research can only truly be achieved by competing at the international level. Participating in centres of excellence supported by regional or international organizations can foster such integration, but only if there is solid commitment from the countries involved.

Third, individual scientists need better security — in jobs, salaries and pensions. Rebuilding links with expatriate scientists and encouraging them to return is vital, as is support for scholarships to study abroad and to bring back the skills required at home. This is one objective of the SESAME training programme, supported by the International Atomic Energy Agency in Vienna.

Finally, science can be an excellent tool for building trust and promoting peace. This is particularly true for large and costly facilities or projects, because such cooperation involves not only scientists but administrators and politicians, sometimes at the highest level. An outstanding example is CERN, which helped to bring together European states after the Second World War. Scientists from not-so-friendly countries work together in solving difficult scientific and technical challenges. They don't just exchange papers, but spend day and night shifts together. In the case of SESAME, Israelis and Palestinians, Turks and Cypriots sit around the same table with colleagues from other countries discussing their common problems peacefully. These are the reasons why organizations such as CERN and SESAME have been established by UNESCO under the slogan "science for peace". There is much to gain, and little to lose, by embracing such ideas. ■

Herwig Schopper is former director-general of CERN and president of the SESAME council.

The golden age

Conditions favouring the ascendancy of Arab culture in the eighth to the thirteenth centuries.

- Patronage of science by persons in high positions. Important examples include the early Abbasid caliphs of Baghdad (AD 754 to 833). Princes and ministers found pleasure or reputation in supporting science. Sultan Ulugh Beg at Samarkand, in modern-day Uzbekistan, even performed his own astronomical observations.
- Support for injunctions in the Koran that it is the bounden duty of every Muslim, man or woman, to acquire knowledge.
- A liberal and tolerant attitude to knowledge from all sources; international contacts and exchange of ideas.
- Good social conditions for scientists. Imam Ghazzali (1058–1111) said: "There are no countries in which it is easier for a scholar to make a provision for his children."

1. Arab Human Development Reports available at <http://cfapp2.undp.org/rbas/ahdr.cfm#>
2. Atta-ur-Rahman & Nasim, A. *Nature* **432**, 273–274 (2004).
3. Maziak, W. *Science* **308**, 1416–1418 (2005).
4. World Development Indicators 2006; available at <http://devdata.worldbank.org/wdi2006/contents/home.htm>
5. Sahel, W. Brain drain threatens future of Arab science *SciDev.Net* (3 June 2004).

See Editorial, page 1; News Features pages, 19–29.

BOOKS & ARTS

Beautiful models

The dynamics of evolutionary processes creates a remarkable picture of life.

Evolutionary Dynamics: Exploring the Equations of Life

by Martin Nowak

Belknap Press: 2006. 384 pp. \$35, £22.95, €32.30

Sean Nee

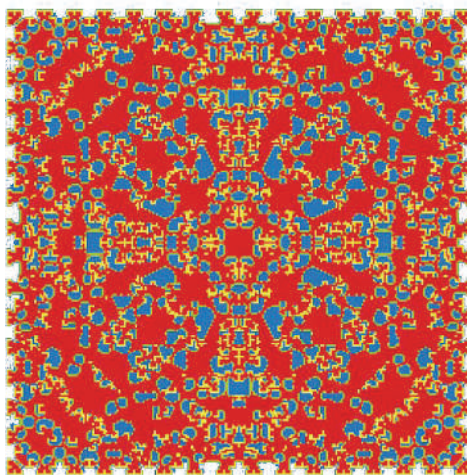
Martin Nowak is undeniably a great artist, working in the medium of mathematical biology. He may be a great scientist as well: time will tell, and readers of this book can form their own preliminary judgement.

In his wanderings through academia's firmament — from Oxford, through Princeton's Institute for Advanced Study to his apotheosis as professor of biology and mathematics at Harvard — Nowak has seemingly effortlessly produced a stream of remarkable theoretical explorations into areas as diverse as the evolution of language, cooperation, cancer and the progression from HIV infection to AIDS. *Evolutionary Dynamics*, based on a course he gives at Harvard, is a comprehensive summary of this work. Although Nowak certainly displays his own oeuvre to great advantage, this book is not purely self-indulgent. His final chapter is an annotated bibliography of other work in the many fields he discusses that is both fair and scholarly: in other words, he cites me.

Many entities replicate. HIV replicates in people's bodies, as do cancer cells. Our genes replicate when we reproduce. Replication may occur with errors as mutation. Natural selection occurs when entities with different properties replicate at different rates, and random chance may also intervene to dilute the action of selection. These are the basic elements of the evolutionary process: if you doubt that such simplicity can produce anything interesting, look around you. Evolutionary dynamics is the mathematical modelling of these processes in a variety of biological scenarios.

A good work of art should stimulate, challenge and, usually, be aesthetically pleasing. Some of Nowak's work in evolutionary dynamics is, literally, visually appealing. But all his work has a beautiful elegance. In time we will see which parts of it become embedded in our way of understanding the various phenomena that inspired him.

Consider, for example, the course of HIV infection. After infection, the virus is initially kept under control by the host's immune system. Over time, mutant virus appears that can escape control by immune cells and multiply



Weaving a spell: Martin Nowak models cooperators and defectors to create patterns like Persian rugs.

until, in turn, this new 'strain' also comes under their control. Nowak's model of the dynamics of this interplay between the virus and the immune system shows a long period during which the virus is under control until a threshold number of strains exist and the immune system collapses. Indeed, the behaviour of the mathematical model elegantly mimics the course of progression from initial infection to AIDS: how could something so beautiful not be true?

For me, the highlight of the book is the chapter on evolutionary graph theory. This is based on a simple reconsideration of the simplest model of evolution, which is that, at successive points in time, an individual in the population dies and is replaced by the progeny of another individual, according to whatever rules of natural selection are being considered. We can visualize this in terms of a graph in which one node can be replaced by a copy of a node connected to it. This is an idea that could have occurred to any of us, but most of us would not have seen how to develop it further. In Nowak's hands, the idea is a springboard: he's off! He designs graphs that amplify, and others that hinder, the efficacy of natural selection compared to the entropic force of random chance — there are bursts, stars, superstars, funnels and metafunnels (see Fig. 3 in *Nature* 433, 312–316; 2005). We get new theorems, such as the isothermal theorem, which tells us what kind of graph can alter the power of natural selection. The chapter fizzles with breathtaking brio. Is the work relevant to anything?

Who knows? Who cares? It's a riot.

Nowak takes the view that ideas in evolutionary biology should be formulated mathematically. An easy retort would be the observation that Darwin managed quite well without mathematics. But, in fact, Darwin did not realize the enormous potential potency of natural selection until he absorbed Thomas Malthus' exposition of the counterintuitive consequences of exponential growth — a fundamentally mathematical insight. Certainly, some ideas that are essentially quantitative must be explored mathematically. But there are plenty of other interesting theoretical areas. Consider genomic imprinting, whereby genes in a fetus are expressed differently depending on whether they come from the father or mother. Nowak's Harvard colleague David Haig has explained this phenomenon

in terms of evolutionary conflicts between parents about investment in the fetus, an explanation that is fascinating, predictive, falsifiable and entirely verbal.

Nowak is much younger and more successful than me. Also, he did not have the modesty to put a question mark after the book's subtitle. So I wanted to hate this book and pen poison to hurt him. I could, for example, chortle that he goes from the most basic model of predator-prey dynamics to Andrey Kolmogorov's eight mathematical conditions for limit cycles in a single page. I could cackle that he assumes that readers know the concept of 'measure' from advanced analysis, and then wonder how many readers he is writing for. But after each mathematical excursion, Nowak provides a perfectly clear and intuitive verbal explanation of what has just happened.

I therefore have no choice but to end positively. This is a unique book. It should be on the shelf of anyone who has, or thinks they might have, an interest in theoretical biology. And if you want to have a punt about what might be considered important new science in the future, this would be a much better buy than another recent book, generously illustrated with pictures of cellular automata but with the much grander aim of revolutionizing science, by another *wunderkind* who also trod the Oxford–Princeton trail.

Sean Nee is at the Institute of Evolutionary Biology, University of Edinburgh, Ashworth Laboratories, King's Buildings, Edinburgh EH9 3JT, UK.

HARVARD UNIV. PRESS

The pork-barrel diet

What to Eat: An Aisle-by-Aisle Guide to Savvy Food Choices and Good Eating
by Marion Nestle

North Point Press: 2006. 624 pp. \$30

Tim Lang

What to Eat is based around a simple idea: how to choose our weekly shopping from the stacked shelves of the average supermarket to achieve a healthy diet. Shopping, like walking or breathing, is deceptively simple. We do it naturally until something happens to make us realize it is both complex and astonishing. In a world where thousands of messages urge us to eat this or drink that, where most supermarkets stock up to 30,000 items, and where value-for-money can drown out value-for-health, choosing the right food to keep our bodies healthy has become very complex.

When Clarence Saunders opened the first self-service grocery store in Memphis, Tennessee, in 1916 he had no idea that persuading consumers to act as shop assistants would instigate a world-wide cultural revolution. Huge savings were achieved, and power relations across the food-supply chain were restructured. Not without reason are the modern supermarket giants such as Wal-Mart, Tesco, Ahold and Carrefour treated with awe in management circles. Their logistics and ruthless supply-chain management keep the shelves full. The humble shopkeepers of old have become today's key gatekeepers of consumerism.

If the restructuring of the food economy is the background to Marion Nestle's excellent new book, the immediate problem she sets out to resolve is how to balance supermarket shopping and good health. Nestle makes a superb guide. A leading US public-health nutritionist with a long and honourable track record of standing up for the public in the sometimes heated world of American food and health politics, she has the advantage of having worked inside government. She was a key player in the US surgeon general's 1988 *Report on Nutrition and Health*, and has seen at first hand the sometimes surreal high-level negotiations behind the lofty aspirations for evidence-based policy — something she explored in her two previous books, *Food Politics* (University of California Press, 2002) and *Safe Food* (University of California Press, 2003).

Like other areas of science, nutrition is framed by political realities. Whose interests are being served? Will this or that power bloc be affected if we tell people to cut down on sugar or meat? In a world where food is very big money, and where nigh on a billion starve while a billion overeat, we are seeing a bizarre globalization of 'pork-barrel politics' in which particular constituencies benefit while

everyone pays. Land that should be producing food for health is producing food for wealth.

This wider context is germane to Nestle's book. How, she asks, can the informed consumer make sense of health demands in an overconsuming, warped food culture? She gives her answers in a walk round a virtual American supermarket, always drawing on real foods and real situations. I imagine her doing this with a trail of enquiring, questioning consumers in a modern socratic food dialogue. "See this food that's been presented as wholesome, good for a healthy heart or endorsed by science? Well, it's not that simple..."

Starting with fruit and vegetables (which usually greet us to remind us we are in a food shop, not just another store in the mall), then progressing to dairy, dairy substitutes, meat, fish, convenience and frozen foods, processed foods, beverages and soft drinks, breads and finally specialties such as baby foods, she stops us and asks: which food would you buy? And then she gives us the data to inform our choice.

On eggs, for example, she unravels the hyperbole around omega-3 fatty acids now being put (via feed) into eggs to counteract their association with cholesterol. After unpicking the complexities of whether high-cholesterol foods automatically lead to high cholesterol in the consumer (it all depends, but they don't help), she boils it down to advice to eat no more than one egg a day, and if you eat that one, avoid other foods made with eggs and other high-cholesterol foods. The book is full of such conclusions: she never spares the intricacies, but

always brings us to a point where we can make choices about whether to buy.

On the complex issue of fish, Nestle teases apart the nutritional profiles according to method of production, and shows how she personally juggles the various demands on the informed consumer. She considers the nutrients of different fish, safety and contamination issues, whether they are farmed or wild, processed or fresh, coloured or natural. Today, even bread is made with fish oils. As she says, "I prefer fish as a source of omega-3s." Me too.

In exasperation at not easily being able to get what any decent consumer would want — health-enhancing, safe, environmentally benign, affordable, trustworthy food — she sometimes reminds us that there can be no individualized solution to this juggling of ethics, science, health and the market. Her answer, given in a purple passage, lies in advocacy to reframe the 'rules of engagement' in the ongoing food war.

Gradually, governments, the wider public and some thinkers at various points in the food chain are realizing that the policy mix is failing. Health education doesn't work. Excess consumer choice is part of the problem. Developing countries are sleepwalking into Western eating lifestyles. Obesity is rampant. Health bills are rocketing, threatening rich as well as poor countries. Our current mode of eating is also ecologically damaging.

Books like this are so important in keeping that bigger picture before the policy-makers. Dare we even think that, one day, the supermarket chains might stop selling unnecessary food? While we are exhorted to eat less, they keep on selling more.

Tim Lang is at the Centre for Food Policy, City University, Northampton Square, London EC10HB, UK.



Asking for help: consumers are faced with an array of difficult choices about which food is healthiest.

S. S. UNAL/CORBIS

A precise past

In its day it served both as a thing of beauty and as a precise, portable calculator and observational tool. The astrolabe was used in astronomy, navigation and surveying, and by Muslims to determine the direction of Mecca and the times of prayer. An important scientific instrument from antiquity to the Renaissance, it was so perfectly designed that it remained in its essentials unchanged for 1,500 years.

Astrolabes at Greenwich, edited by Koenraad van Cleempoel (Oxford University Press, £99.50, \$179.50), explores the collection of more than 50 astrolabes at the National Maritime Museum at Greenwich, London. The engraved Safavid example shown here, from around 1660, exquisitely fuses art and science

B.K.



NATIONAL MARITIME MUSEUM, LONDON

Having faith

Minds and Gods: The Cognitive Foundations of Religion

by Todd Tremlin

Oxford University Press: 2006. 256 pp. \$30, £17.99

Ann Cale Kruger and Melvin Konner

For millennia, religious faith has driven human history, for good or ill. Faith has inspired architectural feats and horrific destruction, soaring music and hateful speech, personal sacrifice and cold-blooded murder. For centuries, historians, philosophers, anthropologists and theologians have wrestled with understanding the mystery and power of religion. Most psychologists have not. But now a small group of thinkers is turning to them for no less than a full scientific explanation of religious beliefs. In *Minds and Gods*, Todd Tremlin bravely attempts to discover the “natural cognitive foundations” of religious thought and, more specifically, seeks a “complete, detailed explanation of the relation of heavenly gods and earthly minds”.

The book is a partial success. More than a third of it explains in detail that human evolution happened, that science has discovered many interesting things about the brain, and that the mind is fundamentally modular. The first two of these conclusions are banal, and the third is more controversial than it seems here. But none of them was needed in such detail to prepare for the book's central thesis.

Tremlin argues that universal human cognitive processes naturally produce gods. In particular, an agency detection device (ADD) and a theory-of-mind module (ToMM) lead us to expect an agent behind every event, and to expect that agent to have a mind. We err on the

side of attributing agency where there isn't any — a trait that no doubt served our ancestors well. And we endow the presumed agent with mental properties that, super-powers notwithstanding, are similar to our own.

The problem is that in reducing complex functions to elemental components, something gets left behind. Tremlin concedes that “ADD and ToMM are not the only mental mechanisms that underpin religious thought... gods evoke intense feelings and emotional experiences”, and argues that “belief in gods indeed is largely an activity of the heart”. Yet he devotes only a few pages to emotion.

The twin human needs for attachment and meaning form the foundation of faith. Unlike other animals, we foresee the distant future, and the vision is unsettling. We ask why we are here and whether life has a purpose; we are anxious about being alone. Religious beliefs may recruit the cognitive mechanisms outlined here, but they are largely fuelled by fear, loneliness and longing. Parents, friends and leaders may abandon or disappoint us, but gods do not. They care about our behaviour and destiny, and know what we do for good and ill, infusing every moment with meaning. Any claim to understand belief must seriously address not just human cognition, but the rich and complex human needs that pervade and energize thinking.

Religion is scarcely uncharted scientific territory. Darwin, in *The Descent of Man*, wrote that religious devotion is highly complex, “consisting of love, complete submission to an exalted and mysterious superior, a strong sense of dependence, fear, reverence, gratitude, hope for the future, and perhaps other elements”. William James (not cited by Tremlin) and

Sigmund Freud (mentioned in passing), the founders of different psychological traditions, both wrote books about the mental processes in religious ideas and practices. James emphasized the inner religious struggle between melancholy and happiness, and pointed to trance as a cognitive mechanism. Freud stressed fear and pain, the need for a powerful parent to care for us, the obsessional nature of ritual, and the hypnotic state a community can induce.

One of us has some religious faith, the other once did. Like most thoughtful people, we wonder where and how faith might be instantiated in mind and brain. There must be a cognitive mechanism that allows a belief in gods, but there must also be more than just conceptual tools to explain the range of behaviours inspired by belief. People routinely believe many things for which there is no evidence, so perhaps it is really science and scepticism that are unnatural and need explaining.

This book represents the emerging field of the cognitive science of religion, and it adds something to our understanding: the natural human penchant for detecting agency and for imagining other minds allow us to conceptualize gods, and thus serve the emotions that motivate religious faith. Perhaps in the future, Tremlin and others working in this new field will attempt to integrate it with older ideas to develop a more comprehensive psychology of religion. Meanwhile, it is good, especially in a world endlessly troubled by religious differences, to have a new theoretical analysis suggesting that there are core religious beliefs that transcend our differences and that stem from universal features of the human mind. ■

Ann Cale Kruger teaches developmental psychology at Georgia State University. Melvin Konner, author of *The Tangled Wing*, teaches anthropology and behavioural biology at Emory University, Atlanta, Georgia 30322, USA.



The opera *Violet Fire* highlights the creativity and the loneliness of Nikola Tesla.

Creativity on the wings of a dove

Violet Fire, an opera about the life of physicist Nikola Tesla, fails to spark.

Horace Freeland Judson

A ballerina, all in white, with fine legs, an odd head-dress and a long floating white train, skitters across the stage, bent double, her elbows up behind her back, stroking the air. Ah! She is a pigeon! She represents — what?

Violet Fire — a hyper-modern multimedia opera about the visionary physicist and electrical engineer Nikola Tesla — had its première at the National Theatre of Belgrade in July and opened in New York in October at the Brooklyn Academy of Music. It prompts a fundamental question: how, in theatre, cinema or opera, can one possibly express the internal aspects of the creative process? Imagine, for a moment, a movie about the poet John Donne. The script would wallow in his disastrous elopement and marriage, and search his anxious face as he preached at St Paul's ("Therefore never send to know for whom the bell tolls: it tolls for thee"). But could it catch him writing a poem? We'd see him sucking the feather of his goose quill and counting out the metre on his knuckles.

In *Violet Fire* (www.violetfireopera.com), the librettist Miriam Seidel, supported by minimalist composer Jon Gibson, pose, but then evade, the problem of the psychic source of scientific and technological innovation. To be sure, minimalist composers often seek out unlikely, difficult operatic subjects — think of Philip Glass's *Einstein on the Beach* or John Adams' *Nixon in China*. Yet Tesla's story is rocky indeed. Born in Croatia in 1856, he died in New York in 1943 — in penury, despite holding more than 700 patents. He

is best known for a series of inventions at the end of the nineteenth century that gave the world alternating current and its chief applications. He patented radio broadcasting years before Marconi. His plans grew ever more ambitious, culminating in a scheme to build a huge tower on Long Island to transmit wireless communication and free, wireless energy worldwide.

But the tower was abandoned when J. P. Morgan withdrew the money. *Violet Fire* begins with a darkened stage, a backdrop with projected schematic representations of the tower, flickering bolts of energy, and Tesla, sung by tenor Scott Murphree, meditating lugubriously. "Perhaps I was a little premature, perhaps too far ahead of time. My tower... sea breeze whistling through the ruins. Perhaps I was a little premature..." — and the long recitative repeats from the start.

Many attempts to convey creativity creatively have been made, and almost always foolishly. The problem is worse in the sciences, because the mental musings, the dialogues in the head, are necessarily more abstract and recondite, as is the finished product. And Tesla worked alone — indeed, at the extreme of loneliness. His inventions sprang from a private, uncanny, intuitive sense of the nature of electrical phenomena coupled with an extraordinary, concentrated visual imagination. In the 1920s, his days of fame long faded, journalists in New York could find him at any time of year sitting on a bench in Central Park, feeding the

pigeons. He had no close collaborators, few friends (although Mark Twain was one), and no intimates. One white dove he considered his companion, with whom he was in communion.

Those pigeons work hard. Seidel and Gibson need them as one of several visual, verbal and musical surrogates for Tesla's creative imagination and for the normal interpersonal sources of operatic tension. Seidel is an American writer who admits she did not study his science and technology; Gibson is a long-time collaborator of Glass.

The second scene finds Tesla on the park bench sprinkling birdseed, while a reporter sings a list of Tesla's chief inventions. "At night and in secret, Nikola Tesla lavishes his love on pigeons." The white pigeon flits through this scene and the rest of the opera, dancing around Tesla at moments of creativity or anguish. Sometimes a small crowd appears to wonder at Tesla's marvels. Minimalist music surely sometimes rises to lyricism; Gibson's is slow, obsessive. Another woman in white, Mirjana Jovanovic, sings the role of the white dove. At the end of the opera she beckons him to an apotheosis atop a scaffolding tower: "Fly into the cloudbank, sink into the dreambank... one hundred million volts... now we are all mourning doves."

The cumulative effect, unfortunately, is a sentimental mystification of the scientific imagination.

Horace Freeland Judson is at 807 West University Parkway, Baltimore, Maryland 21210, USA.

PHOTONICS

A cooling light breeze

Khaled Karrai

Mirrors confine light, and light exerts pressure on mirrors. The combination of these effects can be exploited to cool tiny, flexible mirrors to low temperatures purely through the influence of incident light.

Looking at how a poppy oscillates back and forth in a gentle breeze tells us something about the rigidity of the flower's stem and the strength of the prevailing wind. Observing the movement of a tiny mirror attached to a stem-like post in a 'photon breeze' might be similarly illuminating. Such displacements could reveal the spooky quantum-mechanical behaviour of the mirror itself and maybe even gravity's role in quantum mechanics¹ — no mean feat. The problem is that under normal, room-temperature conditions such effects are masked: the mirror contains an exceedingly large number of thermally excited atoms, so it is in a state of permanent random agitation around its average position on its stalk. And cooling by traditional cryogenic means does not go far enough to remove these 'non-coherent' thermal effects.

In this issue, three papers^{2–4} show two ways of using the pressure of incident photons to damp down thermal agitations. Gigan *et al.* (page 67)² and Arcizet *et al.* (page 71)³ both build on an earlier method^{5,6} for mirror cooling, and for the first time succeed in using radiation pressure to cool miniature mirrors down from room temperature to effective temperatures of just 10–20 kelvin. And Kleckner and Bouwmeester (page 75)⁴ reach a record low temperature of 0.135 K by using a variant of radiation-pressure cooling that uses two lasers in a method known as cold damping⁷. Although spotting intriguing quantum-mechanical effects is probably still some distance away, the research marks a promising step in that direction.

To understand how these experiments work, imagine a conical plunger that is rigidly attached to a mass on a spring (Fig. 1). If such a system is in thermal equilibrium with its environment, it will normally be animated by permanent random (brownian) vibrations around its average position. These vibrations peak at a frequency corresponding to the natural frequency of oscillation of the mass and spring, and can be used to regulate the leakage of light from a cavity that is continuously filled

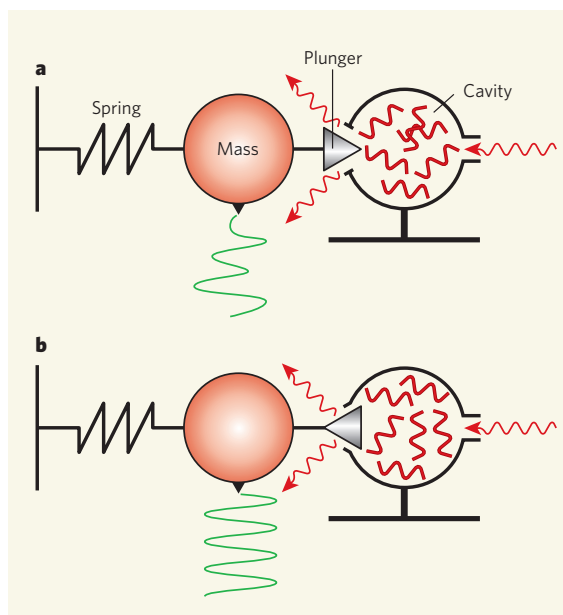


Figure 1 | Calming light. In the radiation-pressure experiments^{2–4}, a conical mirror-like plunger controls light leakage from an optical cavity, and is mounted rigidly on a thermally vibrating mass on a spring. These vibrations move the plunger, and so change the photon density in the cavity and the radiation pressure on the mirror. Owing to light's finite speed, these changes do not occur instantaneously, resulting in 'back-action' in the system. **a**, When the mirror is mounted with its tip pointing into the cavity, the back-action always acts to counter the original thermal fluctuation, resulting in optical cooling of the mirror. **b**, If, instead, the mirror is mounted with its tip pointing out from the cavity, the delayed changes in radiation pressure amplify the fluctuation so that the system enters into a sustained, optically pumped mechanical self-oscillation.

with photons from a separate light source.

The photons exert pressure not just on the walls of the cavity, but also on the plunger, acting to force it outwards. The strength of the pressure depends on the number of photons stored in the cavity at any one time. This is in turn determined by the leakage rate, and so by the position of the plunger. At each sudden random fluctuation of the plunger's position, the cavity fills or empties to a new photon density, and so a new radiation pressure. This readjustment cannot happen instantaneously, as light travels at a finite velocity. While it takes

place, a temporary 'back-action' force results that is crucial to the optical cooling, or quiescence, scheme.

Assume, in the first instance, that the tip of the plunger cone is pointing into the cavity (Fig. 1a). With this geometry, if the initial fluctuation drives the plunger outwards, more photons leak out, and the pressure inside the cavity is reduced. The result is a suction force that acts to pull the plunger back in again. Conversely, when the initial fluctuation acts to drive the plunger inwards, an excess pressure is created in the cavity, resulting in a back-action that pushes the plunger out again. The net result is that the plunger and the attached mass experience a viscous drag that dampens the system's random fluctuations, whatever their initial direction. In other words, the system cools down.

A regular damping mechanism, such as a hydraulic shock absorber or immersing the oscillating body in oil, cannot bring about cooling in this way: the damping fluid would simply heat under the effect of the thermal fluctuations and thus reheat the mass. The beauty of the optical scheme is that, despite the high energy density of photons in the cavity, the light acts as an 'ultracold' damping fluid that does not restore heat to the body. The excess energy gained from the damping will escape irreversibly into the surrounding vacuum, taking away some of the thermal energy that drives the random vibrational motion.

Incidentally, the converse heating effect can be obtained by using a plunger with its conical tip oriented outwards instead (Fig. 1b). In this case, an initial thermal fluctuation outwards will limit the number of photons escaping, increasing the pressure in the cavity, and a fluctuation inwards will lead to a greater leakage of photons, reducing the internal pressure. Thus, whichever way it goes, the thermal fluctuation is amplified.

The experimental challenge facing the authors of these papers^{2–4} is keeping the photons in the cavity over a timescale similar to the natural oscillation time of the mass-spring

system. One can then ensure that the photons impinge on the mirror at just the right time in the system's cycle to achieve a maximum damping (or amplification) effect — rather as a hand on a child's swing at the right time in the cycle increases or decreases its amplitude of oscillation.

With tiny mirrors, this kind of control is far from trivial. Gigan *et al.*² and Arcizet *et al.*³ achieve it by constructing a mirror that forms one end of a very high-quality optical cavity and functions as plunger, mass and spring all in one. Kleckner and Bouwmeester⁴, by contrast, use the cavity only for 'reading out' the position of the fluctuating flexible mirror through precise measurements of the transmitted and reflected light-intensity profiles. The position parameter is then fed with a delay into an electronic feedback loop that controls the intensity of a second laser beam separate from the source of the cavity photons. This beam impinges directly on the mirror⁷. Such a configuration mimics the function of the plunger, but the precise timing of the second beam that thus becomes possible provides a tremendous amplification to the damping, and hence a record optical quiescence effect.

In all these experiments^{2–4}, the degree of cooling achieved so far is limited by the heating that results from vibrations of the mirror's flexible attachments, and most probably from residual

optical absorption by the mirror. To observe the promised quantum-mechanical effects, cooling to just a few millikelvin is needed. That could require the use of microfabrication techniques to produce mechanical oscillators of lower mass that are more easily damped, and cavities of increased optical quality.

Among the prizes for such endeavours could be the chance to study quantum superpositions of a photon and a macroscopic mechanical oscillator. That in turn might find practical use in ultra-precise methods for displacement sensing and for measuring the mass of single atoms and molecules. The road to that destination is a long one; but it is at least now well signposted. ■

Khaled Karrai is in the Center for NanoScience, Ludwig-Maximilians-Universität, Geschwister-Scholl-Platz 1, 80539 München, Germany.
e-mail: karrai@lmu.de

1. Penrose, R. *The Road to Reality* (Vintage, London, 2005).
2. Gigan, S. *et al. Nature* **444**, 67–70 (2006).
3. Arcizet, O., Cohadon, P.-F., Briant, T., Pinard, M. & Heidmann, A. *Nature* **444**, 71–74 (2006).
4. Kleckner, D. & Bouwmeester, D. *Nature* **444**, 75–78 (2006).
5. Braginsky, V. B. & Vyatchanin, S. P. *Phys. Lett. A* **293**, 228–234 (2002).
6. Hühberger-Metzger, C. & Karrai, K. *Nature* **432**, 1002–1005 (2004).
7. Cohadon, P. F., Heidmann, A. & Pinard, M. *Phys. Rev. Lett.* **83**, 3174–3177 (1999).

EVOLUTIONARY BIOLOGY

To work or not to work

David C. Queller

Coercion, not kinship, often determines who acts altruistically in an insect colony. But underlying affinities for kin emerge when coercion is removed: kin selection is what turns suppressed individuals into altruists.

When Shakespeare's Prince Hamlet remarked that his uncle Claudius was "A little more than kin..." he was referring to the added relationship of stepfather that came after Claudius killed Hamlet's father and married his mother. "A little more than kin" could also describe a feature of many social insects. Owing to their odd genetic system, called haplodiploidy, full sisters in ants, bees and wasps are related by 3/4, more than a little above the standard value of 1/2. Historically, this extra kinship figured prominently in the acceptance of W. D. Hamilton's theory of kin selection¹, which holds that workers evolved to altruistically forgo reproduction because they can pass on more of their genes by raising siblings. But a series of papers^{2–5}, including one on page 50 of this issue³, shows that the other half of Hamlet's description — "...and less than kind" — may be more apt. Workers are less than kind both because they must be coerced into their 'altruistic' roles and because workers are

often also the ones doing the coercing.

Of course, kin selection is not just about relatedness; a little more or less kinship can matter less than larger differences in the costs and benefits of altruism^{6,7}. A study of a Malaysian hover wasp by Field *et al.*² provides an elegant demonstration of this point. Some workers work harder than others, and relatedness to the queen does not explain the difference. In this species, rank comes with age. The second-oldest female is the heir apparent, and she reduces her risky foraging. Field and colleagues' experiments showed that this is causal. Removing the second-ranked female causes the third-ranked female to reduce her foraging in line with her improved prospects. Similarly, the amount of prospective gain also mattered; experimentally reducing colony size (adults and brood) caused second-ranked females to increase altruistic foraging, consistent with the diminished value of inheriting the queenship.

Thus, workers slacken off when they may

become queens and work hard when that path is blocked. In social insects that, like the hover wasp, have small colonies and morphologically similar queens and workers, it is usually the queen that does the blocking by her dominance behaviour. For social insects with larger colonies, queen dominance is often replaced by other forms of control. First, there is nutritional coercion. Poorly fed females become small workers and well-fed ones become large queens. This limits the ability of workers to reproduce, but in most species it does not eliminate it fully. Given an opportunity, workers often will lay eggs. In a large colony, the queen could not successfully police all such behaviour and often ignores it. Instead, other workers do the policing, destroying the eggs of their co-workers⁸.

A comparative study of ten policing species by Wenseleers and Ratnieks³ again shows that altruism is modulated more by constraints on worker reproduction than by relatedness. The species with the highest fraction of fully committed altruistic workers — those that do not lay eggs — tend to be those with lower relatedness, contrary to simple expectation. Instead, more workers are fully committed when policing is most effective, as measured by the fraction of worker-laid eggs eaten by either queens or other workers. Workers are not leaping at every opportunity to be altruistic; they are coerced into that role, often by their fellow workers.

Does this mean that Hamilton's kin-selection theory is dead? The answer is no. One could invoke the efforts social insects make to exclude non-relatives from colonies, or the astonishing sex ratios that reflect kin-selected preferences for sisters over brothers⁷. But let us stick with the issue of direct reproduction. For



Figure 1 | Queen control. A comb of the stingless bee *Melipona* with the cell caps removed to show female larvae. Each female receives the same food, and therefore can choose whether to develop as a worker (identifiable here as those with a large head and eyes) or as a queen (small head and eyes). So many females opt to become queens that the workers kill off the surplus³, a form of control that in this species replaces queen selection through nutrition.

T. WENSELEERS

nine of the ten species studied³, there are data on worker reproduction in queenless colonies where there can be no policing in favour of the queen's eggs. Two things change as a result. First, freed from policing, a higher fraction of workers opt to lay eggs. But, remarkably, many individuals still help instead of reproducing, and the fraction that help is now positively correlated with relatedness. Relatedness does matter, and this must be the reason that coercion can induce workers to help.

Wenseleers and Ratnieks⁵ earlier found a similar result when nutritional coercion is absent. In honeybees and their relatives the stingless bees, most species use nutritional coercion to limit queen production to a few at a time. Few are needed, because new queens can reproduce only by usurping the mother queen, or by acquiring a colony on the rare occasions the colony splits into two. The exception is the stingless bee genus *Melipona*. Here, all brood cells are pre-provisioned equally, and then sealed. A developing larva can therefore choose for herself whether to develop as a queen, with a larger abdomen, or as a worker, with larger fore parts (Fig. 1). Many females choose to develop as queens, showing their preference in the absence of constraint. This can result in hundreds of surplus young queens in a colony, so a new level of control has evolved. The workers slaughter the excess queens, so that the

nest evokes the climactic scene of Hamlet, with royal corpses littering the stage. This is a great waste, but shows that it is nutritional coercion that normally keeps queen numbers in check. Yet this case, too, reveals that coercion is not everything and relatedness is important. More than 75% of females still choose the altruistic worker role, and the proportion is higher in species with higher relatedness⁵.

Finally, it should be remembered that Hamilton's kinship theory is not just about altruism *per se*, but about how all traits of altruistic workers evolve. When honeybee ancestors first evolved sociality, the workers could not waggle dance to convey information to each other, or suicidally detach their stings to better repel enemies, or police each other. These features, and all specialized features of workers, had to evolve by kin selection, through their indirect effects on relatives who could pass on genes for these traits. For example, the surprising positive correlation between relatedness and worker laying³, which has been confirmed in a much larger comparative study⁴, is expected under policing. Low relatedness among workers favours workers policing each other. Thus, although policing keeps suppressed workers from fully expressing their kin-related interests, policing is itself kin selected.

Many social conflicts create winners and losers. But only kinship allows evolution to make

creative use of the social losers, turning them into reproductive police, exquisite communicators and heroic defenders. When Hamlet suffered the slings and arrows of outrageous fortune, he debated putting an end to himself. Social insect workers do sometimes choose suicide but, because of kinship, this hamiltonian choice is profoundly different from the hamletian dilemma. The stinging honeybee worker commits suicide when her sting is torn out, but this saves her kin. She is not making an escape from outrageous fortune, but making the best of it — not fearful of what dreams may come, but hopeful for what genes may come. However socially constrained her life may have been, her last action makes her own clear statement: long live the kin!

David C. Queller is in the Department of Ecology and Evolutionary Biology, Rice University, PO Box 1892, Houston, Texas 77251-1892, USA. e-mail: queller@rice.edu

1. Hamilton, W. D. J. *Theor. Biol.* **7**, 1–52 (1964).
2. Field, J., Cronin, A. & Bridge, C. *Nature* **441**, 214–217 (2006).
3. Wenseleers, T. & Ratnieks, F. L. W. *Nature* **444**, 50 (2006).
4. Wenseleers, T. & Ratnieks, F. L. W. *Am. Nat.* www.journals.uchicago.edu/cgi-bin/resolve?id=doi:10.1086/508619
5. Wenseleers, T. & Ratnieks, F. L. W. *Proc. R. Soc. Lond. B* **271**, S310–S312 (2004).
6. West-Eberhard, M. J. *Q. Rev. Biol.* **50**, 1–33 (1975).
7. Queller, D. C. & Strassmann, J. E. *BioScience* **48**, 165–175 (1998).
8. Ratnieks, F. L. W. *Am. Nat.* **132**, 217–236 (1988).

GEOPHYSICS

Same old magnetism

Edward Irving

Latitudes at which ancient salt deposits occur show that Earth's magnetic field has always aligned along its rotation axis. One possible implication is that ancient global glaciations were not caused by a realignment of this axis.

In a paper of admirable scope and thoroughness that appears on page 51 of this issue¹, David Evans analyses the magnetization locked into rocks associated with salts from all over the globe that have been deposited over the past 2,500 million years. Taking as a working model the 'geocentric axial dipole' — the idea that, averaged over thousands of years, the magnetic field at Earth's surface resembles the field of a magnet, or dipole, at Earth's centre^{2,3} — these magnetizations and this model provide clues to the past evolution and interplay of Earth's magnetism, climate and geography.

In the geocentric axial dipole model, the north and south poles of Earth's internal magnet are aligned along Earth's axis of rotation. This simple axial form is thought to be caused by rotational forces that guide the motions of Earth's conducting liquid core, and so constrain the average surface field. Under favourable circumstances, rocks become magnetized along the direction of the ambient geomagnetic

field as they are formed. Thus, by sampling sequences of rocks with formation dates spanning several thousands of years, one can determine the past average direction of the field, the 'palaeolatitude' of the sampling locality and the position of the palaeomagnetic pole at the time. For the past 5 million years, these poles coincide with the present rotational pole; the giant dipole model has therefore been valid for at least this long.

For rocks of much earlier ages, the palaeomagnetic poles determined from rocks from different sampling sites are widely dispersed. This is the result of continental drift and sea-floor spreading in the intervening period. If we restore the continents to their original positions using the geometrical methods of plate tectonics, palaeomagnetic pole positions agree very well⁴. Such corrections go back some 200 million years, and again imply that the geomagnetic field has been a geocentric dipole for that period.

But this evidence does not tell us that the field was also axial. To determine this, one first assumes that the geocentric axial dipole model holds, and determines the latitudes at which temperature-sensitive deposits were laid down from their magnetization directions, or, in the case of salts, those of similarly aged rocks. If these palaeolatitudes are compatible with the modern latitudes of similar deposits, the geocentric axial dipole model is likely to be valid.

The deposits that are the object of Evans's studies¹ are known as evaporites. They comprise beds of, among other things, gypsum, anhydrite, halite and potassium salts, and are of huge economic importance. They were formed by intense evaporation of enclosed saline lakes or sea water. The conditions for their formation must therefore have been hot and dry, as expected typically in latitudes lower than 30°. Very near the Equator, however, it is too wet for them to form.

Evans shows that, consistently over the past 2,500 million years, evaporites have been deposited predominantly between latitudes 10° and 35° (Fig. 1, overleaf). This is a beautifully documented testament to uniformitarianism — the doctrine that today's geological processes have always occurred in a broadly similar manner.

Interest in the interplay between the geomagnetic field and ancient climate zones has been spurred by evidence in a wide range of latitudes, including at sea level near the Equator,



50 YEARS AGO

Not being English or French, but of speech a mid-west American from Manitoba and Dakota, it did not mean anything to me when in 1906 I heard the Eskimos of the Mackenzie Delta and north-eastern Alaska speaking of spruce gum as 'kutsuk'... Years later, perhaps in 1912, I learned that the word for the gum of a Brazilian plant, and of other South American plants, is 'caoutchouc'... Turning now to Greenland, in Sam. Kleinschmidt: "Den Grønlandske Ordbog", I find "kutsuk... Gummi, Campher og lignende producter"... When I first began to talk about this, I was told it was coincidence. But the intellectual climate is changing, among other things in linguistics, and now... [some] think 'connexion' a likelier word than 'coincidence'. Vilhjalmur Stefansson
From *Nature* 3 November 1956.

100 YEARS AGO

Great Bowlers and Fielders. Their Methods at a Glance by G. W. Beldam and C. B. Fry — Following up their interesting volume on "Great Batsmen," the accomplished authors of "Great Bowlers and Fielders" have practically completed all that action photography can teach us regarding the methods of the great cricketers. The present



handsome volume with its 464 action photographs registers for all time the successive positions of the body in the act of bowling of some of the most celebrated bowlers of our day, and also certain very characteristic attitudes of a number of our best fielders... [This] one represents W. Rhodes at the beginning of his final swing, and is chosen partly because of the perfection with which the grip of the ball is indicated.
From *Nature* 1 November 1906.



Figure 1 | Ancient evaporite. This white-coloured sulphate evaporite cliff (about 10 metres high) is interbedded with grey carbonate and mudstone layers within the 800-million-year-old Minto Inlet Formation on Victoria Island in northern Canada. Palaeomagnetic results from age-equivalent rocks in northwestern Canada indicate a latitude of 17° at the time of deposition that is consistent with modern arid climate zones.

that intermittent periods of glaciation covered the whole Earth between 750 million and 550 million years ago. Theories abound as to why this should have occurred, and one proposal⁵ is that Earth's obliquity might have changed drastically at the time. Obliquity is the tilt of Earth's equatorial plane to its ecliptic — the plane of its orbit around the Sun — and is currently 23.5°. If this tilt exceeds around 58°, high latitudes would get more solar heat than low latitudes, and this could account for low-latitude glaciations. The new data¹ prove problematic for such models. Before and after the global glaciations, Evans consistently finds a strong low-latitude, off-Equator peak in evaporite deposition, indicating that obliquity then was not very different from now.

Evans's calculated palaeolatitudes do vary slightly through time, which he divides into four intervals running backwards. These are: interval 1, 250 million years ago (Myr) to the present; interval 2 (370–250 Myr); interval 3 (600–370 Myr); and interval 4 (2,500–600 Myr). The evaporite latitudes do tend to be lower in intervals 4, 3 and 2 than in interval 1, but the changes are irregular. There are two chief explanations for this variation. The first is that before interval 1 the geomagnetic field had long-term axial non-dipole components that endured unchanged for millions of years. These could have perturbed the geomagnetic field sufficiently to cause palaeomagnetic estimates of latitudes based on the geocentric axial

dipole model to be systematically too low by as much as 10°. Alternatively, the field could have remained a simple axial dipole throughout, with the aberrations reflecting, for instance, continental drift or polar wandering.

Evans provides strong evidence for the validity of the geocentric axial dipole model throughout interval 1. Before that, he notes arguments⁶ favouring long-term axial non-dipole fields in interval 2, finds no grounds for them in interval 4, and seems undecided about interval 3.

I will focus on interval 2. Here, Evans accepts the previously advanced view⁶ that Alfred Wegener's grouping of continents into a supercontinent, known as Pangaea A, persisted, not greatly changed, back through the entire interval. There are significant problems with such a long-lived Pangaea. Although there are firm correlations⁸ throughout interval 2 of very thick stratigraphic sequences within Gondwana and within Laurussia (respectively the clustered southern and northern continents that came together to form Pangaea A), there are no comparable correlations between Gondwana and Laurussia. Thus, placing these two supercontinents together in Pangaea A during interval 2 is problematic. Palaeomagnetic data⁷ for late interval 2 based on the assumption of an axial geocentric dipole field place portions of northern Gondwana at the same latitude as southern Laurussia, implying an impossible overlap of the two continents, one on top of the other, of around 1,000 km. But invoking the long-term axial non-dipole components would not remove this overlap⁹, because in such low latitudes their effect would be small and equal in both places. It is worth noting, too, that Evans's analysis does not explicitly require long-term axial non-dipole components in interval 2.

Evans makes no steady overall commitment to either of the contending explanations for the variations in the evaporite palaeolatitudes. I hope I will be forgiven for doing so. It seems reasonable to me to take the estimates of evaporite latitudes at face value and accept that, during the past 2,500 million years — apart from a relatively brief global glaciation at the end of interval 4 — hot and dry climates were typical of tropical, but not equatorial, latitudes, and that long-term non-dipole components have always been small or negligible. In other words, one should accept the geocentric axial dipole model as demonstrably successful for interval 1 and use it to settle questions of palaeogeography in earlier intervals. For example, displacing Gondwana to the east during interval 2 by around 3,500 km, a palaeogeography known as Pangaea B, reconciles the palaeomagnetic data with a feasible configuration of the early continents⁹.

Such palaeogeographic solutions are more generally testable than solutions involving long-term non-dipole fields: whereas the geomagnetic dynamo in early ages is an awfully remote item, ancient strata can be directly studied and the phenomena they reflect described. Most

R. H. RAINBIRD

researchers accept interval 2 palaeogeography as settled. I disagree: on the principle of working from the known to the unknown, it could be just the issue that needs further critical consideration before we can confidently approach the question of earlier climate zones. Data of the extensive nature supplied by Evans¹ are what is needed to inform such debates.

Edward Irving is an emeritus scientist at the Pacific Geoscience Centre, Geological Survey of Canada, Department of Natural Resources, PO Box 6000, Sidney, British Columbia V8L 4B2, Canada.

e-mail: tirving@pgc-gsc.nrcan.gc.ca

1. Evans, D. A. D. *Nature* **444**, 51–55 (2006).
2. Gilbert, W. (transl. Mottelay, P. F.) *De Magnete* (1600) (Dover, Mineola, NY, 1958).
3. Hospers, J. J. *Geol.* **63**, 59–74 (1955).
4. Besse, J. & Courtillot, V. *J. Geophys. Res.* **107**, doi:10.1029/2001JB000050 (2002).
5. Williams, G. E. *Earth Sci. Rev.* **34**, 1–45 (1993).
6. Torsvik, T. H. & Van der Voo, R. *Geophys. J. Int.* **151**, 771–794 (2002).
7. Wegener, A. (transl. Skerl, J. G. A.) *Origin of Continents and Oceans* (Methuen, London, 1924).
8. Du Toit, A. L. *Our Wandering Continents* (Oliver & Boyd, Edinburgh, 1937).
9. Muttoni, G. et al. *Earth. Planet. Sci. Lett.* **215**, 379–394 (2003).

CANCER BIOLOGY

Second step to retinal tumours

Valerie A. Wallace

The mutations that cause retinoblastoma are well known, but how they enable the cancer to evade controls on cell division was unclear. Secondary mutations affecting a growth-regulatory pathway have now been identified.

Most cancers arise through a two-step process in which an initiating mutation requires further tumour-promoting mutations to instigate the full-blown disease. Retinoblastoma is a childhood cancer of the retina that is one of the few tumour types for which the initiating genetic lesion is known — both copies of the retinoblastoma (Rb) tumour-suppressor gene are inactivated. However, few of the additional tumour-promoting lesions have been identified. On page 61 of this issue, Laurie *et al.*¹ report that amplification of the numbers of MDMX and MDM2 genes occurs frequently in human retinoblastoma. They show that increased expression of MDMX can promote tumour development in genetic models of retinoblastoma in mice. They also show that a drug called nutlin-3, which blocks some actions of MDM proteins, can stop tumour progression in the eye, opening up a promising avenue for the potential treatment of these tumours.

MDMX and MDM2 are structurally related proteins that act as antagonists of a cell-signalling pathway named after its most famous member — the tumour suppressor p53 (ref. 2). The p53 protein is a gene regulatory factor that normally inhibits cell proliferation and induces cell death in response to cellular stress. The Rb tumour-suppressor pathway also inhibits cell proliferation, in part by blocking the expression of genes required for the cell to divide. Normally, loss of Rb leads to induction of p14^{ARF}, a key activator of p53 (ref. 3; Fig. 1). But in many cancers, p53 activity is blocked by mutations in the p53 gene or alterations to other genes in the pathway, allowing cells to escape death and leading to uncontrolled division. In human retinoblastoma, however, the p53 pathway is intact⁴, and acute Rb loss in

the human retina induces p14^{ARF} expression, which should unleash the lethal potential of p53 in response to this mutation¹. So how do retinoblastoma cells subvert the intact p53 pathway to prevent it from killing them?

Laurie *et al.* provide evidence that the p53 pathway is circumvented in retinoblastoma cells

by increased expression of MDMX or MDM2. These proteins interact with p53, blocking its gene-regulatory activity and promoting its degradation² (Fig. 1). Analysis of human retinoblastoma reveals that the number of MDMX and MDM2 genes is amplified in 65% and 10% of the tumours, respectively, and that this correlates negatively with p53 levels. Moreover, the levels of MDMX RNA and protein were increased in several recently removed human tumours, and the authors confirm that MDMX antagonizes p53-mediated activation of certain target genes, cell-cycle arrest and programmed cell death in retinoblastoma cell lines.

To investigate the functional significance of the MDMX protein in the development of retinoblastoma, the authors turned to mouse models and cultured human fetal retina tissue (retinal explants). In the mouse, the development of retinoblastoma requires the elimination of Rb and one additional Rb family member, either p130 or p107 (ref. 5). However, as occurs in humans, the tumours in these models can develop with intact p53. Laurie *et al.* show that MDMX expression inhibits cell death, promotes proliferation and exacerbates retinoblastoma progression in mice lacking Rb and p107. MDMX expression also blocks cell death and promotes proliferation in human fetal retinal explants where Rb expression was reduced experimentally. This effect was not observed with an MDMX mutant that does not bind to p53. Furthermore, in both the mouse and the human retinal models, MDMX

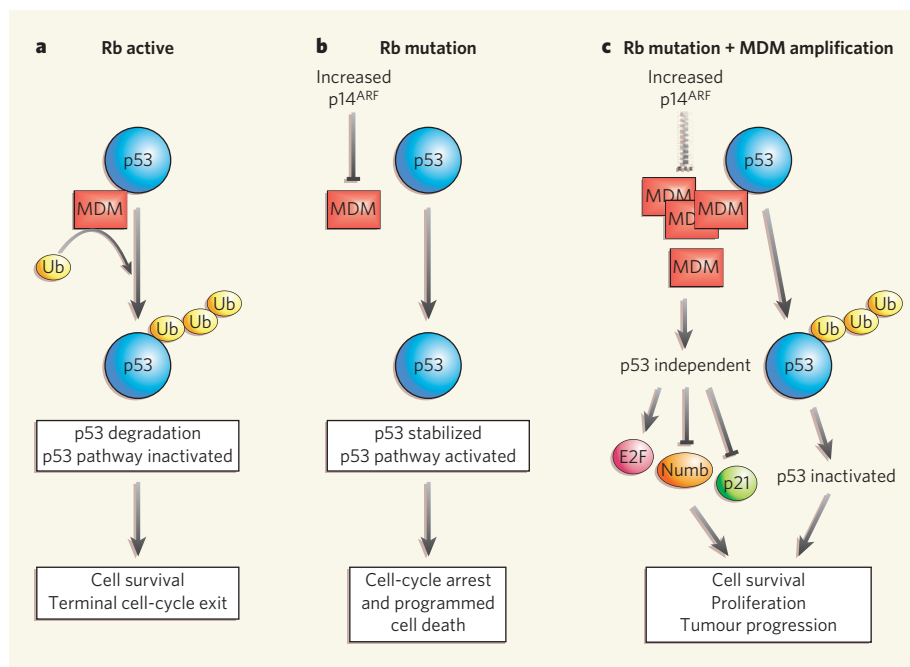


Figure 1 | Role of MDM proteins in retinoblastoma. **a**, Under normal conditions, levels of p53 protein are kept low, partly through negative regulation by MDM proteins. MDM2 is an enzyme that tags p53 with a ubiquitin molecule (Ub), thereby promoting p53 degradation. MDMX also interacts physically with p53 and inhibits its gene-regulatory activity. **b**, Mutation of the retinoblastoma gene (Rb) results in increased production of the p14^{ARF} protein, which in turn leads to inactivation of MDM2, thus promoting p53 pathway activation. This leads to cell-cycle arrest or cell death, or facilitates DNA repair. **c**, Increased expression of MDM proteins in the absence of Rb blocks activation of p53, leading to survival of the abnormal cells and tumour progression. Independently of p53, interactions of MDM with other proteins that regulate cell division, survival and differentiation could also promote tumour progression¹³.

induced a rapid reduction in the expression of cell markers of differentiation — the cellular process of maturation and specialization. This suggests that MDMX may be promoting tumours with a less differentiated and more aggressive character.

On the basis of their findings, Laurie *et al.*¹ conclude that MDMX promotes retinoblastoma by blocking p53 activity. However, this interpretation is inconsistent with previous reports^{6,7} showing that cell death in the retinas of mice lacking either Rb or both Rb and p107 is independent of p53. Moreover, inhibition of programmed cell death does not seem to be an essential feature of retinoblastoma in the mouse⁸. Given these issues, is it possible that the anti-differentiation effects of MDMX are more significant in tumour development than its anti-death effects? Clearly, MDMX in this context can promote tumour development, but as suggested elsewhere⁹, this could be mediated in a p53-independent fashion through its interaction with other key regulatory proteins¹⁰. Interestingly, the growth-promoting and anti-differentiation interactions of MDM2 with other proteins, such as Numb, also require the p53-binding site of MDM2 (Fig. 1)¹¹.

Finally, Laurie *et al.*¹ show that nutlin-3, a drug that blocks the interaction between MDM2 and p53, can unleash p53-dependent death in MDMX-expressing retinoblastoma cells; this has also been shown by Ellison *et al.*¹². Moreover, combined direct delivery to the eye of nutlin-3 and topotecan, a drug that promotes p53 activation, synergistically inhibits tumour growth of intraocularly injected retinoblastoma cells in mice, without being toxic to the animals. However, these studies will need to be extended to models of spontaneous MDMX-driven tumour development. So, although the mechanisms by which MDM proteins facilitate the progression of retinoblastoma require further analysis, their functional relevance in the cancer provides an exciting potential target for its treatment.

Valerie A. Wallace is in the Molecular Medicine and Vision Research Programs, Ottawa Health Research Institute, 501 Smyth Road, Ottawa, Ontario K1H 8L6, Canada, and in the Departments of Ophthalmology and of Biochemistry, Microbiology and Immunology, University of Ottawa.
e-mail: vwallace@ohri.ca

- Laurie, N. A. *et al.* *Nature* **444**, 61–66 (2006).
- Marine, J.-C. & Jochemsen, A. G. *Cell Cycle* **3**, 900–904 (2004).
- Lowe, S. W. & Sherr, C. J. *Curr. Opin. Genet. Dev.* **13**, 77–83 (2003).
- Nork, T. M. *et al.* *Arch. Ophthalmol.* **115**, 213–219 (1997).
- Dyer, M. A. & Bremner, R. *Nature Rev. Cancer* **5**, 91–101 (2005).
- Robanus-Maandag, E. *et al.* *Genes Dev.* **12**, 1599–1609 (1998).
- MacPherson, D. *et al.* *Genes Dev.* **18**, 1681–1694 (2004).
- Chen, D. *et al.* *Cancer Cell* **5**, 539–551 (2004).
- Pacal, M. & Bremner, R. *Curr. Mol. Med.* **6**, 759–781 (2006).
- Zhang, Z. *et al.* *Ann. NY Acad. Sci.* **1058**, 205–214 (2005).
- Juven-Gershon, T. *et al.* *Mol. Cell. Biol.* **18**, 3974–3982 (1998).
- Elison, J. R., Cobrinik, D., Claros, N., Abrams, D. H. & Lee, T. C. *Arch. Ophthalmol.* **124**, 1269–1275 (2006).
- Ganguli, G. & Wslylyk, B. *Mol. Cancer Res.* **1**, 1027–1035 (2003).

PHYSICAL CHEMISTRY

Porous solids get organized

Rutger A. van Santen

A powerful combination of analytical techniques is used to shed light on the complex crystallizations of porous solids. Molecular recognition creates the seeds of order from which complex lattices grow.

Although the era of empirical chemistry is seemingly long gone, many chemical processes are still poorly understood. The formation of crystalline porous solids with nanoscale cavities is a case in point. These materials are extremely useful catalysts for many industrial chemical reactions, but we have little understanding of how they crystallize from their disordered precursors. Such insight could enable better catalysts to be designed. A novel combination of spectroscopic tools has now been used to unravel the mechanistic details of the synthesis of complex nanoporous solids, as reported in two papers in the *Journal of the American Chemical Society*^{1,2}. The results suggest that molecular complexes in the synthetic mixture ‘recognize’ template molecules, so forming clusters from which ordered structures may grow.

Zeolites are porous solids that contain channels up to 1 nanometre long. The building-

blocks of these solids are tetrahedra of oxygen atoms, with a cation at the centre of each tetrahedron. Three-dimensional networks are formed through corner-sharing of the tetrahedra. From the infinite number of theoretical frameworks that can be constructed from these building-blocks, only about 150 different zeolites are found experimentally.

Archetypal zeolites have a lattice framework that is electrostatically neutral, such as is seen in silicalite, which is constructed from SiO₄ tetrahedra. Similar structures, known as aluminophosphates, can be constructed from combinations of aluminium-based tetrahedra (AlO₄) and phosphorus-based tetrahedra (PO₄) (ref. 3). A good example of this is the zeolite known as AlPO₄-5, which contains one-dimensional channels.

Zeolites become catalysts only when reactive complexes are incorporated into the structure. This can be done in two ways. First, one can

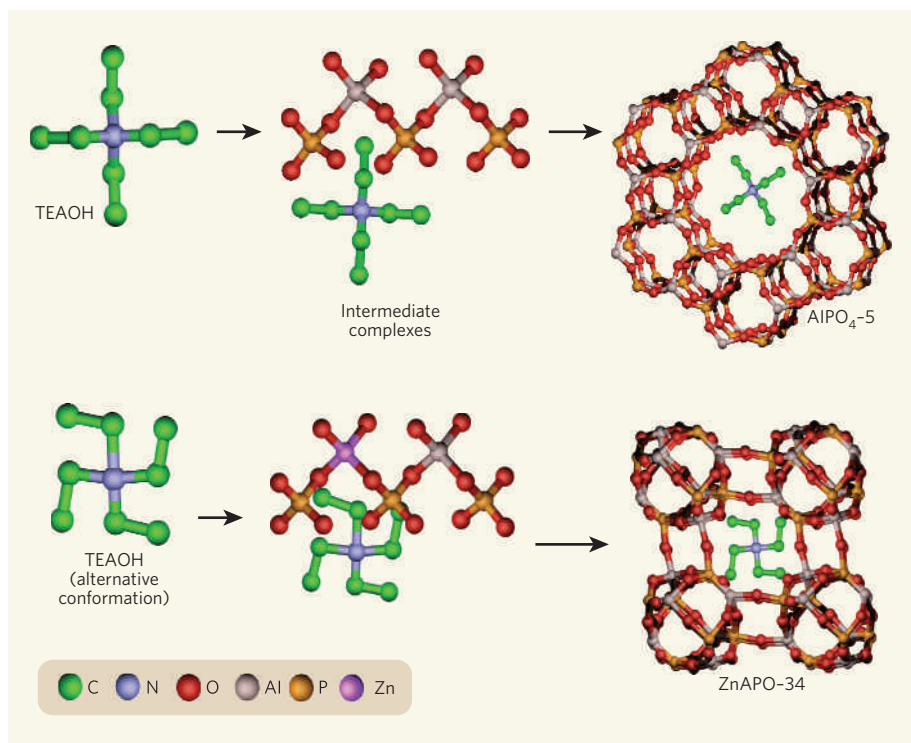


Figure 1 | Template-directed crystallization of microporous solids. **a**, On synthesis, aluminophosphate crystallizes as a porous solid known as AlPO₄-5, which contains regular arrays of microscopic channels. This process is directed by an organic base (tetraethylammonium hydroxide; TEAOH) in the reaction mixture. Weckhuysen and colleagues' data^{1,2} suggest that TEAOH forms a complex with the aluminophosphate molecules that acts as a template for constructing the AlPO₄-5 lattice. **b**, When zinc is incorporated into some of the aluminophosphate molecules, TEAOH may adopt an alternative conformation. This yields a different intermediate complex that acts as a template for the formation of a structure known as ZnAPO-34. (Structures courtesy of Andrew M. Beale.)

replace cations in the lattice framework with cations that have a lower positive charge. For example, when cobalt ions (Co^{2+}) substitute for some of the aluminium ions (Al^{3+}) in aluminophosphates, a porous solid is generated that catalyses oxidations⁴. Such cation substitutions result in a framework that has an overall negative charge. This can be balanced by placing cations in the nanoporous channels. If reactive cations are used, this is a second way to activate catalytic behaviour.

A breakthrough in zeolite science came with the discovery that previously unknown structures could be made by adding organic bases to the chemical reactions used to prepare zeolites⁵. These organic bases seem to act as a template for the formation of micropores and are now commonly used in zeolite synthesis, although the molecular details of their role are poorly understood.

Now, Weckhuysen and colleagues^{1,2} have advanced our understanding of the molecular mechanism for the organic-base-mediated synthesis of zeolites. Zeolites usually form as gels, which then crystallize into the desired microporous solids. The authors used a combination of *in situ* techniques to examine this crystallization process. They applied X-ray scattering methods¹ previously also used to study silicalite synthesis from solution⁶ to probe the dimensions of particle aggregates and the presence of crystals. They combined this with spectroscopy² to investigate the local environment of the atoms.

Weckhuysen and colleagues compared the formation of the chargeless $\text{AlPO}_4\text{-5}$ framework with the negatively charged framework (known as ZnAPO-34) that is formed by replacing aluminium ions in $\text{AlPO}_4\text{-5}$ with zinc ions (Zn^{2+}). The authors followed not only the structural changes in the aluminophosphate gel in real time, but also the conformational features of the organic base (tetraethylammonium hydroxide) used as a template for the crystallization. The tetraethylammonium cation adopts one of two conformations once it takes up its position in the aluminophosphate nanopores.

The authors prepared the ZnAPO-34 structure, which contains spherical cavities rather than channels, simply by adding zinc ions to the aluminophosphate gel. They also made similar structures by adding either cobalt ions (Co^{2+}) or manganese ions (Mn^{2+}) to the gel², showing that the influence of dipositive cations on the lattice topology is general. Their studies revealed previously unknown details of the crystallization process — for example, particles of crystalline material, about 11.5 nanometres in size, initially form in the ZnAPO-34 gel before increasing in size¹. Most intriguingly, not only do ZnO_4 tetrahedra form in the precursor gel before crystallization begins, but the tetraethylammonium ion also adopts the conformation that it will ultimately assume in the crystal. In the case where no zinc is present and $\text{AlPO}_4\text{-5}$ forms, the organic template

takes on the alternative conformation.

These results^{1,2} are highly relevant to the debate on the mechanism of zeolite formation and the role of organic base molecules⁶⁻⁹. The implication is that the tetraethylammonium ion forms a complex with developing zeolite subunits in the gel (Fig. 1), adopting a molecular structure close to that found in the final crystal². This molecular recognition process determines which type of crystal lattice is formed. For ZnAPO-34 , the distinctive spherical cavities of the crystal may actually form first in the gel, forcing the tetraethylammonium ion into the conformation seen in the solid. The near absence of this conformation in the alternative $\text{AlPO}_4\text{-5}$ system corroborates this theory. Furthermore, Weckhuysen and colleagues' data are in line with proposals on the mechanism of silicalite formation from its synthesis solution^{6,9}. It is thought that precursor nanoparticles of silicalite might appear before crystallization, and that these nanoparticles are aggregates of complexes formed between SiO_4 units and the organic base used in the reaction.

Weckhuysen and colleagues^{1,2} have thus shed light on the mechanism of zeolite formation, providing experimental support for the idea that molecular organization occurs before

crystallization. The insights obtained from this work will certainly benefit the search for the next generation of microporous materials. More generally, the authors' powerful combination of analytical techniques will enable the physical chemistry of many other synthetic processes to be monitored *in situ*, so that the structural changes involved can be observed as they happen. ■

Rutger A. van Santen is at the Schuit Institute of Catalysis, Eindhoven University of Technology, Box 513, 5600 MB Eindhoven, The Netherlands. e-mail: r.a.v.santen@tue.nl

1. Beale, A. M. *et al.* *J. Am. Chem. Soc.* **128**, 12386–12387 (2006).
2. O'Brien, M. G., Beale, A. M., Catlow, C. R. A. & Weckhuysen, B. M. *J. Am. Chem. Soc.* **128**, 11744–11745 (2006).
3. Flanagan, E. M., Lok, B. M., Patton, R. L. & Wilson, S. T. in *New Developments in Zeolite Science and Technology* Vol. 28 (eds Murakami, Y., Iijima, A. & Ward, J. W.) 103–112 (Elsevier, Amsterdam, 1986).
4. Raja, R., Sankar, G. & Thomas, J. M. *J. Am. Chem. Soc.* **123**, 8153–8154 (2001).
5. Barrer, R. M. & Denny, P. J. *J. Chem. Soc.* 971–982 (1961).
6. de Moor, P.-P. E. A. *et al.* *Chem. Eur. J.* **5**, 2083–2088 (1999).
7. Davis, T. M. *et al.* *Nature Mater.* **5**, 400–408 (2006).
8. Nokedakis, V., Kokkoli, E., Tirell, M., Tsapatsis, M. & Vlachos, M. *Chem. Mater.* **12**, 845–853 (2000).
9. Kirschhock, C. E. A., Ravishankar, R., van Looveren, L., Jacobs, P. A. & Martens, J. A. *J. Phys. Chem. B* **103**, 4972–4978 (1999).

NEUROBIOLOGY

Crossed circuits

Andrew B. Schwartz

Can the brain be induced to reroute neural information? Such an achievement is crucial if the function of damaged brain areas is to be taken on elsewhere. A study in monkeys explores this prospect.

The rapidly growing field of neural engineering has led to the development of electronic devices that interact directly with neurons, with the aim of examining fundamental neural operations and of replacing damaged brain functions. In work described on page 56 of this issue, Jackson, Mavoori and Fetzi¹ use a self-contained electronic circuit implanted in the brains of monkeys to demonstrate a basic feature of learning — the ability to change the routing of neural information — that could prove useful in rehabilitative therapy.

The authors used off-the-shelf components to build what they call a Neurochip. Neurons in the brain's motor cortex fire impulses during voluntary movements, and the Neurochip picks up these impulses, or 'spikes', by means of a recording microelectrode placed near the neurons. Through the Neurochip, the spikes trigger an electric-stimulus pulse through a second, stimulating electrode at an adjacent site in the motor cortex. Because the circuit is self-contained, this spike–stimulus sequence could be carried out continuously for 24 hours while

the animal went about its normal activities in its cage (termed the conditioning period). The idea behind the Neurochip was that, by picking up the signals from one neuronal circuit and transferring them to another repeatedly over a 24-hour period, a new information route might be formed. This would have obvious potential with regard to rehabilitation after neural injury, but it would also give insight into learning, which is thought to occur as neural circuits and connections are strengthened through use.

The authors also used a technique called intracortical microstimulation (ICMS), which provides a measure of neural connectivity^{2,3} between the cortical neurons and muscles, to assess whether the neural circuit had been rerouted by the Neurochip during the conditioning period. In ICMS, high-frequency bursts — or trains — of electrical pulses (in the range 10–100 microamps) are passed through microelectrodes in the cortex, activating a large network of neuronal connections spanning the cortex, subcortical structures and the spinal cord to generate muscle twitches.

Cortical neurons that project to different muscles are in fact intermixed in the part of the motor cortex that is likely to be activated by ICMS⁴. Because of the extent of this network, it is not possible to delineate a specific functional route between the stimulation site and the observed muscle twitch using ICMS by itself, and it is difficult to ascribe a specific causal role to any cortical neuron. But although ICMS is only a rough measure of connectivity, it can indicate when the pathways change in response to external stimuli. For instance, activation of the muscles by ICMS changes rapidly with peripheral-nerve injury⁵, or even when a limb is placed in a splint for just a few hours⁶.

Fetz and colleagues¹ used ICMS before the conditioning period to define the direction of force produced at the wrist when they stimulated the motor cortex around the Neurochip recording and stimulation sites (Fig. 1). A combination of muscles was probably activated from each ICMS site, and different combinations produced different force directions. ICMS applied at the recording site generated torque in one direction, whereas that applied at the stimulating site generated torque in another. The Neurochip was then used to apply single shocks at the stimulation site each time a spike occurred at the recording site during the conditioning period.

When ICMS was applied to the recording site after this conditioning period, the torque direction produced at the wrist had changed, becoming more like that produced by the stimulation site (a shift from flexion towards radial deviation). So it seems that a change may have occurred in the network between the recording site and the muscle. The authors suggest that connectivity is enhanced between the recording and stimulation sites, and that this is facilitated by horizontal connections in the motor cortex, but that is just one possibility. ICMS induces many neurons to fire in relation to one another, in ways that may not happen during normal behaviour. But the authors also used an alternative technique that measures the correlation of spiking between pairs of neurons or between a

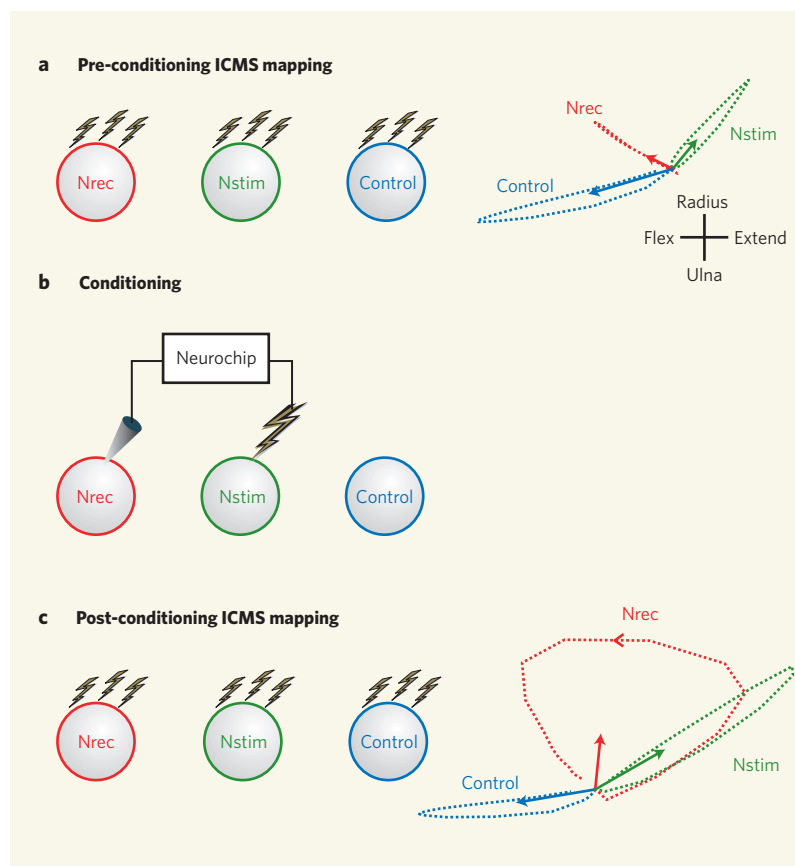


Figure 1 | Reorganizing a neural circuit. Fetz *et al.*¹ implanted a 'Neurochip', which had a recording microelectrode (Nrec) and a stimulating microelectrode (Nstim), in the motor cortex of a monkey. **a**, Using a technique called intracortical microstimulation (ICMS), they passed a series of electric pulses into the cortex next to the two Neurochip electrodes and at a control region to map the muscle responses at the monkey's wrist. The right panel shows the wrist torque produced by ICMS, with the arrows showing the mean trajectory. **b**, The next step was a conditioning period, when the Neurochip recorded the activity of a single neuron in the cortex. Every time there was an activity spike in that neuron, the Neurochip sent an electrical pulse through the stimulating microelectrode into another neuron. **c**, After 24 hours of conditioning, the ICMS response was measured again. The response produced by ICMS near the Neurochip recording site had changed towards that produced near the Neurochip stimulation site, showing that the neural circuit had been partially rerouted by conditioning.

neuron and muscle, avoiding the complication of artificial connectivity. So far, these data show a clear correlation only between activity in the single neuron being recorded and the two opposing antagonist muscles in the wrist. If the horizontal-connectivity hypothesis is correct, then further studies should show that ongoing spike activity at the two electrode sites is more correlated after conditioning. In addition, spikes recorded at the recording site should show an increased cross-correlation to the activity patterns (EMG) of muscles that are correlated to the stimulation-site activity.

The results show clearly that stimulation at a particular cortical location activates pathways to multiple muscles, and that the conditioning procedure may change the effective bias of these pathways. Whether this is facilitated by local connections between neurons in the cortex, or through more remote projections to, for example, subcortical and spinal structures, is still an open question. The predominant

pathways descending to the musculature from the motor cortex have many neuronal junctions, with nerve fibres projecting to other cortical regions, the thalamus, subcortical structures, and throughout the intermediate layers of the spinal cord. Even the few fibres that end directly on motor units in the spinal cord facilitate contraction in a variety of muscles that generate torque in different directions. The pairing strategy used by Fetz and colleagues¹ is likely to alter the connections throughout this extensive network, changing the muscle bias elicited from the ICMS site. This explanation is supported by the finding¹ that the 20-millisecond delay of the stimulus following the conditioning trigger spike has an optimal effect (merely changing horizontal connections in the motor cortex would generally require a conduction delay of only about 2 milliseconds).

Nonetheless, these effects observed by Fetz and colleagues are specific and long-lasting (persisting for more than a week after conditioning), warranting further experiments using similar technology in primates. Such neural devices, and the new re-

search they spawn, will help to elucidate how the brain can change and adapt within the complexity of an intact nervous system, and will undoubtedly change our view of how learning occurs in higher-order species, including our own.

Andrew B. Schwartz is in the Departments of Neurobiology, Bioengineering and Physical Medicine and Rehabilitation, University of Pittsburgh School of Medicine, E1440 BST, 200 Lothrop Street, Pittsburgh, Pennsylvania 15213, USA.
e-mail: abs21+@pitt.edu

1. Jackson, A., Mavoori, J. & Fetz, E. E. *Nature* **444**, 56–60 (2006).
2. Asanuma, H. & Sakata, H. *J. Neurophysiol.* **30**, 35–54 (1967).
3. Stoney, S. D., Thompson, W. D. & Asanuma, H. *J. Neurophysiol.* **31**, 659–669 (1968).
4. Rathelot, J. A. & Strick, P. L. *Proc. Natl Acad. Sci. USA* **103**, 8257–8262 (2006).
5. Donoghue, J. P., Suner, S. & Sanes, J. N. *Exp. Brain Res.* **79**, 492–503 (1990).
6. Sanes, J. N., Wang, J. & Donoghue, J. P. *Cereb. Cortex* **2**, 141–152 (1992).

MATERIALS SCIENCE

Qubits in the pink

Pieter Kok and Brendon W. Lovett

Crystal imperfections known as nitrogen–vacancy defects give some diamonds a characteristic pink colour. Appropriately manipulated, these defects might have rosy prospects as the ‘qubits’ of a quantum computer.

According to materials scientist F. C. Franck, “crystals are like people; it is only the defects that make them interesting”. Ronald Hanson and colleagues would probably agree: writing in *Physical Review Letters*¹, they report new developments in the study of negatively charged ‘nitrogen–vacancy defects’ in diamond. These systems are rapidly becoming a front-runner for use as the basic unit of quantum information — the ‘qubit’ — in a solid-state quantum computer.

The lattice of carbon atoms that makes up diamond can contain various substitutional impurities, such as nitrogen or boron atoms. These defects give diamonds their colour, and are often called colour centres. Nitrogen–vacancy (NV) defects give diamond a pink hue, and arise when a nitrogen atom replaces a carbon atom at a position in the diamond lattice next to a vacant site. It might seem unlikely that these two defects should sit right next to each other, but if the diamond is heated, vacancies can diffuse through the lattice until they encounter nitrogen atoms. When this happens, the ‘random walk’ comes to a halt because the configuration of the two adjacent defects is extremely stable.

NV centres come in two flavours: electrically neutral (NV⁰) and negatively charged (NV[−]). Here, we are interested only in the NV[−] centres, which have an extra electron that is probably donated by another nitrogen defect. The total number of electrons in an NV[−] defect that do not form bonds between neighbouring carbon atoms is six, and these electrons have a ground state with total spin $S = 1$. This spin can have three different orientations, described by a magnetic quantum number $m_s = +1, 0$ or -1 . If the spin tends to take one of these values, it is said to be polarized.

A well-defined, ‘coherent’ spin state can be preserved in the NV[−] centre for a long time (more than 50 microseconds, even at room temperature²). The energy difference between the $m_s = 0$ state and the $m_s = \pm 1$ states is about 12 microelectronvolts. Photons with a microwave frequency of 3 gigahertz have precisely this energy, and can therefore be used to manipulate the spin state of the NV[−] centre. To ‘read out’ the spin state, transitions to much higher energy states are used. These transitions can be induced by shining a laser on the

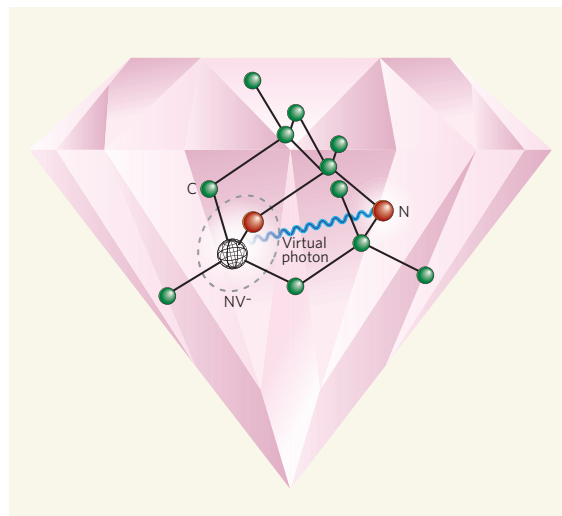


Figure 1 | Defective couple. Hanson and colleagues’ experiments¹ look at nitrogen–vacancy (NV[−]) defects in the proximity of nitrogen (N) defects in the lattice of carbon atoms that makes up diamond. By tuning the energy of the spin states of the defects using a magnetic field, the defects can be brought into resonance, exchanging energy and polarization via so-called virtual photons. This interaction allows spin control of the nitrogen defect using optical control of the NV[−] defect, and the combined system forms a promising two-qubit structure for quantum computing.

diamond. The light that is emitted when the NV[−] centre relaxes back to the lower energy levels tells us what the spin state was. This process is called photoluminescence³.

The NV[−] spins give us much of what we need for a practical qubit — long coherence times and precise external control — and have been widely studied in the context of quantum computing⁴. But considerable obstacles must be overcome before effective quantum computing with NV[−] is possible. Perhaps the toughest of these is scaling the system up to many qubits.

Hanson *et al.*¹ take a significant step towards solving this problem. They demonstrate a coupling between an NV[−] centre and a substitutional nitrogen centre, N, with no associated vacancy (this is not the same nitrogen defect that supplies the NV[−] with its extra electron). This ‘N centre’ also has a spin, but with different magnetic quantum numbers: $m_s = \pm 1/2$. When a magnetic field is applied, the energy-level structure of both the NV[−] and N spins changes, a phenomenon known as the Zeeman effect. At certain values of the field, the energy gaps between the spin levels in the two defects become equal. If there is an interaction between NV[−] and N, this ‘resonance’

allows the two defects to exchange energy and polarization through a so-called virtual photon (Fig. 1). This reduces the photoluminescence signal because the NV[−] is no longer in the correct quantum state to emit light. It is this effect that Hanson *et al.* observe¹.

In a further experiment, the authors bring a microwave field into resonance with the NV[−] defect. This too leads to an exchange of polarization, this time between the NV[−] spin and the field. If the NV[−] is isolated in the diamond lattice, a single dip should be observed in the photoluminescence signal. But Hanson *et al.* see two dips in their signal¹, again because of the interaction with the N centre. The two possible orientations of the N spin cause different effective magnetic fields at the NV[−] centre, and shift the resonance accordingly. Importantly, this shift allows us to read out the spin state of the N centre by simply observing the photoluminescence of the NV[−] centre. The same experiment shows that a specific state of the N centre can be prepared on demand.

As N centres are very common in samples containing NV[−] defects, the interactions described above are a main cause of NV[−] qubits losing coherence. Being able to control an N centre is a promising way of reducing this decoherence to an acceptable level.

The N–NV[−] system can embody two electron-spin qubits, so these experiments represent a significant step towards a larger-scale quantum computer based on the NV[−] system. The authors suggest that several NV[−] centres could be connected through chains of N defects. This seems to us very ambitious;

a better way of achieving a many-qubit register could be to use a form of ‘measurement-based’ quantum computing⁵, in which quantum correlations are created between qubits before any quantum algorithms are executed. In this context, the N–NV[−] system could be ideal for a protocol that creates the required quantum correlations between ‘broker’ qubits through optical manipulations and subsequently transfers them to ‘client’ qubits that have longer decoherence times, but no optical transitions⁶. With such a strategy, the future for quantum computing would look rosy indeed. ■

Pieter Kok and Brendon W. Lovett are in the Department of Materials, University of Oxford, Parks Road, Oxford OX1 3PH, UK.
e-mail: pieter.kok@materials.oxford.ac.uk

1. Hanson, R., Mendoza, F. M., Epstein, R. J. & Awschalom, D. D. *Phys. Rev. Lett.* **97**, 087601 (2006).
2. Kennedy, T. A., Colton, J. S., Butler, J. E., Linares, R. C. & Doering, P. J. *Appl. Phys. Lett.* **83**, 4190 (2003).
3. Jelezko, F., Gaebel, T., Popa, I., Gruber, A. & Wrachtrup, J. *Phys. Rev. Lett.* **92**, 076401 (2004).
4. Santori, C. *et al.* preprint available at www.arxiv.org/quant-ph/0607147 (2006).
5. Raussendorf, R. & Briegel, H. J. *Phys. Rev. Lett.* **86**, 5188 (2001).
6. Benjamin, S. C., Browne, D. E., Fitzsimons, J. & Morton, J. J. L. *New J. Phys.* **8**, 141 (2006).

BRIEF COMMUNICATIONS

Enforced altruism in insect societies

Cooperation among workers and their seeming altruism result from strict policing by nestmates.

Workers of many species of ant, bee and wasp do not lay eggs, despite having functional ovaries¹, but the selective causes of this extreme form of altruism are unclear^{2–7}. Here we show that workers forego reproduction in response to the threat of their eggs being killed, or 'policed', by nestmates. Our results indicate that social coercion helps to explain worker altruism and cooperation in modern-day insect societies^{3–5}.

Why, in some species, do most workers forego direct reproduction? One possibility is that worker altruism is voluntary: in this scenario, high genetic relatedness should drive the evolution of altruism^{2–7} and worker sterility⁷ because higher relatedness increases the indirect benefit of working. Theoretically, however, worker altruism could also be 'enforced' and may have evolved in response to social sanctions^{4–6}. In many species, worker-laid eggs are killed by the queen or by other workers^{1,8,9} and, if these sanctions are effective, the advantage to workers of laying eggs is reduced. As a result, more would be selected to work altruistically, rather than to lay eggs⁶.

The role of sanctions in promoting worker sterility has long been suspected^{8,9}, but has never been tested in a comparative study. We therefore studied ten single-queen species, nine Vespidae wasps (for example, see Fig. 1) and the honey-



Figure 1 | The wasp *Vespa crabro*.

bee *Apis mellifera*. These are the only species for which there are data to quantify both the proportion of non-egg-laying workers, a measure of worker altruism, and two key predictor variables: relatedness among workers, which depends on the frequency of queen mating, and the effectiveness with which worker-laid eggs are policed by nestmates^{1,8,9}. We analysed the data by using individual species as data points and by phylogenetically independent contrasts (PICs) to control for phylogenetic non-independence. (See supplementary information.)

Figure 2 shows that, as predicted, fewer workers reproduce when the effectiveness of policing worker-laid eggs is higher ($P = 0.00004$, Fig. 2a; using PICs: $P = 0.000006$, see supplementary information). This supports the hypothesis that worker altruism is enforced. Contrary to the voluntary-altruism hypothesis⁷, however, higher relatedness does not lead to increased altruism. In fact, the reverse is true — a larger proportion of the workers reproduce in species where relatedness is high ($P = 0.004$, Fig. 2b; using PICs: $P = 0.04$; see supplementary information). However, this is predicted by policing theory, because low relatedness more strongly selects for workers to police each others' reproduction^{1,8}.

Our results also show that policing effective-

ness is negatively correlated with relatedness ($R = -0.60$, one-tailed $P = 0.03$; using PICs: $F(1,8) = 4.69$, one-tailed $P = 0.03$); this contrasts with our results from queenless colonies, in which the relationship is reversed and higher relatedness results in a smaller proportion of the workers laying eggs ($R = -0.79$, $P = 0.007$; using PICs: $P = 0.03$; see supplementary information). This is as expected from theory, as in queenless colonies policing does not occur and its inhibitory effect is lost⁷. However, the effect of relatedness in promoting altruism remains⁷.

The key role of relatedness in the evolution of self-sacrificing behaviour is widely recognized^{2,5}. The origin of insect societies is one of the most cited examples, and high relatedness was probably required for worker behaviour first to evolve^{2,5}. Nevertheless, our results show that in modern-day insect societies it is mainly social sanctions that reduce the numbers of workers that act selfishly. In this, they provide evidence for something that has proved notoriously hard to demonstrate in human society: that better law enforcement can lead to fewer individuals behaving antisocially¹⁰.

Tom Wenseleers*†, Francis L. W. Ratnieks*‡

*Institute for Advanced Study, 14193 Berlin, Germany

†Present addresses: Zoological Institute, Catholic University of Leuven, 3000 Leuven, Belgium (T.W.); Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK (F.L.W.R.)

e-mail: tom.wenseleers@bio.kuleuven.be

- Wenseleers, T. & Ratnieks, F. L. W. *Am. Nat.* **168** www.journals.uchicago.edu/cgi-bin/resolve?id=doi:10.1086/508619.
- Hamilton, W. D. *J. Theor. Biol.* **7**, 1–52 (1964).
- Alexander, R. D. *Annu. Rev. Ecol. Systemat.* **5**, 325–383 (1974).
- Frank, S. A. *Evolution* **57**, 693–705 (2003).
- Lehmann, L. & Keller, L. *J. Evol. Biol.* **19**, 1365–1376 (2006).
- Wenseleers, T., Hart, A. G. & Ratnieks, F. L. W. *Am. Nat.* **164**, E154–E167 (2004).
- Wenseleers, T., Helanterä, H., Hart, A. G. & Ratnieks, F. L. W. *J. Evol. Biol.* **17**, 1035–1047 (2004).
- Ratnieks, F. L. W. *Am. Nat.* **132**, 217–236 (1988).
- Ratnieks, F. L. W. & Visscher, P. K. *Nature* **342**, 796–797 (1989).
- Levitt, S. D. *Am. Econ. Rev.* **87**, 270–290 (1997).

Supplementary information accompanies this communication on Nature's website.

Received 22 June; accepted 14 September 2006.

Competing financial interests: declared none.

doi:10.1038/44450a

BRIEF COMMUNICATIONS ARISING online
 ▶ www.nature.com/bca see Nature contents.

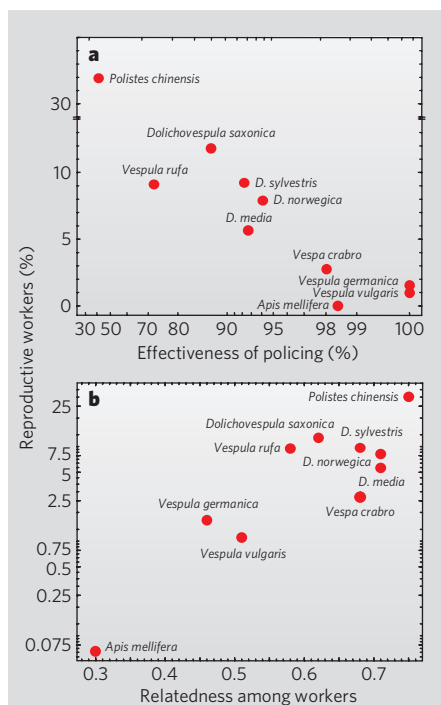


Figure 2 | Effect of sanctions and relatedness on worker altruism in social insects. **a**, If altruism is enforced, more workers should remain sterile when their reproduction is more effectively policed by nestmates, which is what occurs ($R = -0.94$, $P = 0.00004$; effectiveness of policing is reverse $\log_{10}$ -transformed). **b**, If altruism is voluntary, greater altruism and less worker reproduction should be seen when relatedness is high, but the opposite occurs ($R = 0.82$, $P = 0.004$; percentage of reproductive workers is $\log_{10}$ -transformed). The effectiveness of policing is defined as the probability of worker-laid eggs being killed relative to queen-laid eggs; reproductive workers are shown as the percentage of workers with active ovaries (see supplementary information).

GEOCHEMISTRY

Does U–Pb date Earth's core formation?

Arising from: B. J. Wood & A. N. Halliday *Nature* **437**, 1345–1348 (2005)

Constraining the timing of the formation of Earth's core, which defines the birth of our planet, is essential for understanding the early evolution of Earth-like planets. Wood and Halliday¹ and Halliday² discuss the apparent discrepancy between the U–Pb (60–80 Myr) and Hf–W clocks (30 Myr) in determining the timescale of Earth's accretion and core formation. We find that the information the authors present is at times contradictory (for example, compare Fig. 1 in ref. 1 with Fig. 1 in ref. 2) and confusing and could suggest that the U–Pb clock constrains core formation better than the Hf–W system. Here we point out the limitations of the U–Pb system and show that the U–Pb age cannot be used to argue for protracted accretion and/or core formation (>50 Myr) because this clock only records the processes that occurred during the last 1% of Earth's accretion and core formation in the Wood and Halliday mechanism¹.

For both the U–Pb and Hf–W systems to be able to date Earth's accretion and core formation, we need accurate estimates of the following parameters: the initial Solar System tungsten and lead isotope ratios; the bulk Earth elemental ratios of Hf/W and U/Pb; the present tungsten and lead isotope composition of the silicate Earth; and the present U/Pb and Hf/W ratios of the silicate portion of Earth. Whereas we know all the required parameters for the Hf/W system, several of the parameters for the U/Pb system are highly uncertain and are unlikely ever to be determined accurately enough to infer age resolution of a few tens of millions of years at 4.5 Gyr ago.

First, the U/Pb ratio of the bulk Earth (that is, μ_{TOE} in ref. 1) is not well known, as it is a ratio of a refractory and a volatile element. Rocky planetary objects are depleted in volatile elements, including Pb to various degrees. We therefore have no way of knowing *a priori*

what the building-blocks of Earth are made of in terms of the U/Pb ratio. The same is not true for the Hf–W system, as both elements are refractory. The bulk Earth Hf/W ratio has been estimated from that of primitive chondritic values with a high degree of confidence^{3,4}. Using refractory elements for estimating the bulk composition of planets is one of the cornerstones of modern geochemistry.

Second, because of the long half-lives of ^{238}U (~4.5 Gyr) and ^{235}U (~0.7 Gyr), the lead isotope composition of terrestrial rocks varies widely and is continuously evolving. As a result, there is no easy way of deriving a single value for the present lead isotope composition of the bulk silicate Earth that can be used to date early Earth events precisely within the first 30 Myr. The same is not true for the Hf–W system. The now-extinct radionuclide ^{182}Hf only has a 9-Myr half-life and decayed to ^{182}W . Every rock on the surface of Earth records a single value of ^{182}W , which is 2 parts per 10,000 higher than the bulk Earth value (obtained from chondrites). The difference is a consequence of metal extraction during core formation, which also produces a high Hf/W ratio in the silicate portion of Earth. Thus, we have a well-determined tungsten isotope composition for the bulk silicate Earth, but a highly uncertain one for lead isotopes.

The accretion process is not a well-defined point in time and, in fact, is technically still continuing today. The best designation of the 'end' of our planet's formation (>99% mass accreted) is probably the final major event in Earth's accretion process, the formation of the Moon by a Mars-sized impactor. This time is constrained by the tungsten isotope data to be about 30 million years after the formation of the Solar System^{4–8}. New tungsten isotope data from lunar samples reinforce this conclusion (see Fig. 3b of ref. 9).

Wood and Halliday¹ invoke the late-stage sulphide phase segregation into the core to explain the inconsistency between U–Pb and Hf–W. However, to account for the increased U/Pb ratio in the silicate Earth by lead partitioning into the core, note that additional mass added to the core after the last Moon-forming giant impact is trivial (<1%)¹. The U–Pb clock therefore sees one of the 'trees but not the forest' of core formation, as the U–Pb record necessarily misses >99% of the core-formation processes. The U–Pb system is strongly influenced by Earth's surface processes, and the so-called 'Pb paradox' still has many alternative interpretations to core formation¹⁰. Late sulphide segregation may explain the putative age difference between U–Pb and Hf–W, but it is misleading to portray the U–Pb ages as the time of Earth's core formation¹.

Qing-zhu Yin*, Stein B. Jacobsen†

*Department of Geology, University of California, Davis, California 95616, USA

e-mail: yin@geology.ucdavis.edu

†Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA

1. Wood, B. J. & Halliday, A. N. *Nature* **437**, 1345–1348 (2005).
2. Halliday, A. N. *Nature* **427**, 505–509 (2004).
3. Harper, C. L. & Jacobsen, S. B. *Geochim. Cosmochim. Acta* **60**, 1131–1153 (1996).
4. Yin, Q.-Z. et al. *Nature* **418**, 949–952 (2002).
5. Schoenberg, R., Kamber, B., Collerson, K. D. & Eugster, O. *Geochim. Cosmochim. Acta* **66**, 3151–3160 (2002).
6. Kleine, T., Münker, C., Mezger, K. & Palme, H. *Nature* **418**, 952–955 (2002).
7. Jacobsen, S. B. *Science* **300**, 1513–1514 (2003).
8. Jacobsen, S. B. *Annu. Rev. Earth Planet. Sci.* **33**, 531–570 (2005).
9. Kleine, T., Palme, H., Mezger, K. & Halliday, A. N. *Science* **310**, 1671–1674 (2005).
10. Hofmann, A. W. in *The Mantle and Core: Treatise on Geochemistry* Vol. 2 (ed. Carlson, R. W.) 61–101 (Elsevier-Pergamon, Oxford, 2003).

doi:10.1038/nature05358

GEOCHEMISTRY

How well can Pb isotopes date core formation?

Arising from: B. J. Wood & A. N. Halliday *Nature* **437**, 1345–1348 (2005)

Timescale and the physics of planetary core formation are essential constraints for models of Earth's accretion and early differentiation. Wood and Halliday¹ use the apparent mismatch in core-formation dates determined from tungsten (W) and lead (Pb) chrono-

meters to argue for a two-stage core formation, involving an early phase of metal segregation followed by a protracted episode of sulphide melt addition. However, we show here that crust–mantle Pb isotope systematics do not require diachronous core formation. Our

observations indicate that very early (≤ 35 Myr) core formation and planet accretion remain the most plausible scenario.

Core formation could have entailed a severe fractionation of the U/Pb ratio in the mantle, ultimately causing the average lead (radiogenic

daughter) isotope composition of the silicate Earth to be displaced from the array defined by undifferentiated meteorites. The extent of this displacement would depend on the time of fractionation. Therefore, the lead isotope composition of the bulk silicate Earth (BSE), if it were precisely known, could serve to estimate the timing of core formation²; however, it is not well known. This is because homogenization as a result of mantle stirring has not been sufficiently fast to eradicate the isotopic heterogeneities that reflect differences in time-integrated U/Pb ratios between parts of the silicate Earth. Attempts to obtain a precise mass balance of the heterogeneous mantle and crustal reservoirs are bedevilled by the unknown masses of reservoirs with characteristic lead isotope compositions.

Specifically, terrestrial lead isotope evolution models have been presented for the crust–mantle system³. In the preferred solution for these, a timescale for core segregation of 60 Myr after Solar System formation, based on the then (1997) prevailing interpretation of chondritic tungsten isotope data⁴ (later shown to be erroneous^{5,6}), was used as an imposed constraint³. This yielded the BSE lead isotope composition estimate shown as KT 97 in Fig. 3 of Wood and Halliday¹. Using this value to derive a core-formation age is therefore circular. The core-formation timescale is not a

critical parameter for terrestrial Pb modelling (Fig. 9 of ref. 3). The much faster core formation now implied by tungsten-isotope models^{5–7} is equally consistent with the lead data.

The BSE lead-isotope estimate MKC 03 (in Fig. 3 of Wood and Halliday¹) was a by-product of finding a refined solution to the first lead-isotope paradox⁸. It depends on the core-formation age preferred by Murphy *et al.*⁸, who maintain that the slope of the calculated 'geochron' corresponds to an age of 4,503 Myr and that this implies catastrophic core formation within about 65 Myr. Wood and Halliday's model¹ returns exactly that age for their proposed stage of lead loss with a late sulphide melt. Again, the argument is circular.

These examples show that it is not possible to determine the lead isotope composition of BSE directly with sufficient precision to constrain a relatively short time interval at the beginning of Earth's history. Early work on the 'U–Pb age' of Earth⁹, which required implausibly protracted core formation (200 Myr), historically influenced all estimates for the lead isotope composition of BSE. Six of the 11 estimates used by Wood and Halliday¹ were published before 1990 and yield an average time of core formation of 91 Myr. The five more recent estimates return a faster average core-formation date of 61 Myr. This demonstrates that, as we learn to appreciate the complexity of terres-

trial differentiation¹⁰, our estimates of the BSE lead isotope composition are moving closer to the meteorite isochron than was previously thought possible⁹. It may be that the true BSE lead composition is compatible with tungsten isotope constraints for very rapid (that is, ≤ 30 Myr) accretion and core formation. We therefore find no solid observational evidence for the complex core-formation process proposed by Wood and Halliday¹.

Balz S. Kamber*, Jan D. Kramers†

*Department of Earth Sciences, Laurentian University, Sudbury, Ontario P31 6B5, Canada
e-mail: bkamber@laurentian.ca

†Department of Geology, University of Bern, 3012 Bern, Switzerland

1. Wood, B. J. & Halliday, A. N. *Nature* **437**, 1345–1348 (2005).
2. Halliday, A. N. *Nature* **427**, 505–509 (2004).
3. Kramers, J. D. & Tolstikhin, I. N. *Chem. Geol.* **139**, 75–110 (1997).
4. Lee, D.-C. & Halliday, A. N. *Nature* **378**, 771–774 (1995).
5. Yin, Q.-Z. *et al.* *Nature* **418**, 949–952 (2002).
6. Schoenberg, R., Kamber, B. S., Collerson, K. D. & Eugster, O. *Geochim. Cosmochim. Acta* **66**, 3151–3160 (2002).
7. Jacobsen, S. B. *Annu. Rev. Earth Planet. Sci.* **33**, 531–570 (2005).
8. Murphy, D. T., Kamber, B. S. & Collerson, K. D. *J. Petrol.* **44**, 39–53 (2003).
9. Allègre, C. J. *Earth Planet. Sci. Lett.* **5**, 261–269 (1969).
10. Rudnick, R. L., Barth, M., Horn, I. & McDonough, W. F. *Science* **287**, 278–281 (2000).

doi:10.1038/nature05359 (2006)

GEOCHEMISTRY

Wood & Halliday reply

Replying to: Q.-Z. Yin & S. B. Jacobsen *Nature* **443**, doi: 10.1038/nature05358 (2006) ; B. S. Kamber & J. D. Kramers *Nature* **443**, doi: 10.1038/nature05359 (2006)

Yin and Jacobsen¹ and Kamber and Kramers² highlight several issues to do with the tungsten (W) and lead (Pb) isotopic chronometry of terrestrial accretion and core segregation, although none of these affect our model or conclusions³. Yin and Jacobsen¹ suggest that it is not clear from our analysis³ that ^{182}Hf – ^{182}W provides a reliable system of chronometry for the bulk of core formation, whereas $^{235/238}\text{U}$ – $^{207/206}\text{Pb}$ relates to the very last stage of sulphide segregation —which was exactly our point³ and is clarified here.

The concern of Yin and Jacobsen¹ that we used different versions of Fig. 1^{3,4} to illustrate scenarios for the accretion of the Earth is addressed in the respective figure captions and texts^{3,4}, which demonstrate the wide range of different kinds of accretion model that underlie the W and Pb isotope modelling. There is no inconsistency. None of the accretion curves shown is likely to represent the actual growth of the Earth, which is underconstrained. ^{182}Hf – ^{182}W chronometry does not provide a unique mass accretion and core-formation history of the Earth^{4,5}.

The difficulties both sets of authors raise^{1,2}

about the determination of the average Pb isotopic composition of the bulk silicate Earth (BSE) are correct, but these are well known and illustrated by the range of existing estimates^{4–6}. They have not prevented Pb isotope estimates being made and inferences drawn about their significance^{6,7}. Explanations for certain facets of the Pb paradox that are based on melting and recycling of material within the BSE are irrelevant; they cannot change the average composition. Either all 11 of these estimates, and the Pb isotope space between them, are incorrect, or the timing and mechanism of U/Pb and Hf/W fractionation are different, with $^{235/238}\text{U}$ – $^{207/206}\text{Pb}$ always yielding longer timescales^{3–7}. This is not a random scatter that reflects a poorer ability to determine the average Pb isotopic composition of the BSE. Rather, there is a systematic offset that provides evidence for a different fractionating process, or a systematic error, as we state^{3,4}. Yin and Jacobsen¹ agree with this¹; Kamber and Kramers² do not, but do not explain how this discrepancy can be eliminated.

The weak constraints on the μ_{TOTE} of the Earth are well known^{4–7}. We used³ an esti-

mate⁷ based on chondrite volatile depletion trends⁴. Estimates based on half-mass condensation temperature (see ref. 6, for example) are slightly higher⁸, but these increase the discrepancy between the timing deduced from $^{235/238}\text{U}$ – $^{207/206}\text{Pb}$ and ^{182}Hf – ^{182}W (ref. 4). Our arguments are, if anything, conservative.

Kamber and Kramers² point out that ^{182}Hf – ^{182}W constraints have been used to determine the average Pb isotopic composition of the BSE^{9,10}. It would have been better not to use their estimates because of the circularity involved. Based on partitioning, the approach of using ^{182}Hf – ^{182}W to predict the Pb isotopic composition of the BSE is flawed³. Our concern was mainly to demonstrate that all estimates were consistent with our model. The three most recent estimates attributable to Kamber and Kramers are similar to the average of the eight earlier, more varied, estimates (see ref. 4). The means of the two-stage model ages are 67 and 81 Myr, respectively, whereas the equivalent calculation for ^{182}Hf – ^{182}W yields about 30 Myr (ref. 11).

Although Kamber and Kramers² clarify how they made their estimates and Yin and Jacobsen¹

reiterate the importance of the ^{182}Hf – ^{182}W system, this does not change any major aspect of our paper. In closing, we point out two inaccuracies in their comments.

Yin and Jacobsen¹ claim that the age of the Moon is well constrained by ^{182}Hf – ^{182}W to about 30 Myr, citing six papers, none of which contains an analysis of a lunar sample and one of which preceded the first data. The ^{182}Hf – ^{182}W age of the Moon is 30 to 55 Myr^{12–14}. If it were earlier, the W isotopic effects would be greater; if it were later, there would be no radiogenic effects. A more precise age will not be determined until the initial and average W isotopic compositions of the Moon, which were derived from an unknown mixture of the silicate Earth and the silicate and metal reservoirs of Theia, are known. An estimate of 40–50 Myr has been used^{3,4}, based on the least radiogenic lunar tungsten to be found so far^{4,5,12–14}; a similar estimate of 30 to 50 Myr was based on the apparent radiogenic ingrowth within lunar reservoirs¹⁴. But there is no basis for stating that it is about 30 Myr.

Kamber and Kramers² claim that the time of core formation was previously constrained to about 60 Myr (ref. 9) on the basis of W isotopic data¹⁵. This is incorrect, because at that time no difference in W isotopic composition between chondrites and the BSE had been resolved. Even now that accurate chondrite data have been acquired, the meaning of the difference between chondrites and the BSE needs to be interpreted in terms of other constraints, such as Pb isotopes. If not, the W isotopic difference could be interpreted in terms of half the core forming at the start of the Solar System and the other half during the Battle of Hastings.

B. J. Wood*, **A. N. Halliday†**

*Department of Earth and Planetary Sciences, Macquarie University, New South Wales 2109, Australia

†Department of Earth Sciences, University of Oxford, Parks Road, Oxford OX1 3PR, UK
e-mail: bwood@els.mq.edu.au

1. Yin, Q.-Z. & Jacobsen, S. B. *Nature* **444**, doi:10.1038/

nature05358 (2006).

2. Kamber, B. S. & Kramers, J. D. *Nature* **444**, doi:10.1038/nature05359 (2006).
3. Wood, B. J. & Halliday, A. N. *Nature* **437**, 1345–1348 (2005).
4. Halliday, A. N. *Nature* **427**, 505–509 (2004).
5. Halliday, A. N. in *Meteorites, Comets and Planets* Vol. 1 (ed. Davis, A. M.) 509–557 (Elsevier–Pergamon, Oxford, 2003).
6. Galer, S. J. G. & Goldstein, S. L. in *Earth Processes: Reading the Isotope Code* (eds Basu, A. & Hart, S.) 75–98 (AGU, Washington DC, 1996).
7. Allègre, C. J., Manhès, G. & Göpel, C. *Geochim. Cosmochim. Acta* **59**, 1445–1456 (1995).
8. Lodders, K. *Astrophys. J.* **591**, 1220–1247 (2003).
9. Kramers, J. D. & Tolstikhin, I. N. *Chem. Geol.* **139**, 75–110 (1997).
10. Murphy, D. T., Kamber, B. S. & Collerson, K. D. *J. Petrol.* **44**, 39–53 (2003).
11. Yin, Q.-Z. et al. *Nature* **418**, 949–952 (2002).
12. Lee, D.-C., Halliday, A. N., Snyder, G. A. & Taylor, L. A. *Science* **278**, 1098–1103 (1997).
13. Lee, D.-C., Halliday, A. N., Leya, I., Wieler, R. & Wiechert, U. *Earth Planet. Sci. Lett.* **198**, 267–274 (2002).
14. Kleine, T., Palme, H., Mezger, K. & Halliday, A. N. *Science* **310**, 1671–1674 (2005).
15. Lee, D.-C. & Halliday, A. N. *Nature* **378**, 771–774 (1995).

doi:10.1038/nature05360

Proterozoic low orbital obliquity and axial-dipolar geomagnetic field from evaporite palaeolatitudes

David A. D. Evans¹

Palaeomagnetism of climatically sensitive sedimentary rock types, such as glacial deposits and evaporites, can test the uniformitarianism of ancient geomagnetic fields and palaeoclimate zones. Proterozoic glacial deposits laid down in near-equatorial palaeomagnetic latitudes can be explained by 'snowball Earth' episodes, high orbital obliquity or markedly non-uniformitarian geomagnetic fields. Here I present a global palaeomagnetic compilation of the Earth's entire basin-scale evaporite record. Magnetic inclinations are consistent with low orbital obliquity and a geocentric-axial-dipole magnetic field for most of the past two billion years, and the snowball Earth hypothesis accordingly remains the most viable model for low-latitude Proterozoic ice ages. Efforts to reconstruct Proterozoic supercontinents are strengthened by this demonstration of a consistently axial and dipolar geomagnetic reference frame, which itself implies stability of geodynamo processes on billion-year timescales.

The principle of uniformitarianism, in which modern geological processes guide our conception of the ancient world, has faced a challenge in the field of Precambrian palaeoclimate. In contrast with Ordovician and younger ice ages, which are characterized by expectedly high palaeolatitudes, Proterozoic glaciogenic deposits have until now yielded purely low to moderate palaeomagnetic latitudes, including several near-equatorial results of high reliability^{1,2}. A simple interpretation of these results invokes the encroachment of ice sheets across all latitudes, a model known as snowball Earth^{3–5}. The concept of a 'hard' snowball with global ice cover enduring for millions of years has been controversial, with some scientists advocating partly unfrozen tropical oceans^{6,7}.

An alternative to global refrigeration invokes a high Precambrian planetary obliquity of more than 54°, which would reverse zonal mean-annual temperature gradients, preferentially spawning ice ages in the tropics rather than near the poles^{8,9}. A single direct measure of Precambrian obliquity by using stromatolite heliotropism¹⁰ suggests a 'normal' low value at about 630 Myr, but this isolated example should be confirmed by similar studies elsewhere. Geophysical considerations on the feasibility of the high-obliquity hypothesis are generally negative^{11–13}, yet the spatial distribution of Precambrian glacial deposits, which lack any well-documented high-latitude examples, currently permit both the snowball and high-obliquity models^{1,2}.

A non-uniformitarian Precambrian geomagnetic field may also be invoked to explain the low palaeomagnetic inclinations. Database-wide palaeomagnetic compilations indicate a shallow bias relative to random sampling of a geocentric-axial-dipole (GAD) field across the spherical surface, which may be explained by the preferential motion of continents into low latitudes, or subsidiary higher-order components to the ancient geomagnetic field¹⁴. Some support for the latter possibility was raised by means of numerical geodynamo modelling with spatially non-uniform core–mantle boundary conditions^{15,16}.

The distribution of ancient evaporite deposits can help distinguish between these non-uniformitarian alternatives. Modern tropospheric Hadley–Ferrel circulation descends on the subtropics of

both hemispheres, where evaporation exceeds precipitation in today's desert belts. Fifty years ago, the emerging science of palaeomagnetism admirably succeeded in reconstructing consistent palaeoclimatic zones in the context of post-Pangaean continental drift^{17–19}. Subsequent studies considered evaporitic rocks as old as the Cambrian period and found deposits concentrated in palaeolatitude bands consistent with modern evaporite belts of 15–35° latitude^{20–23}. Here I revisit the evaporite palaeolatitude test according to the current palaeomagnetic data set and extend it back through Proterozoic time, to the oldest recognized examples of large evaporite basins on Earth.

Evaporite palaeolatitudes

The world's largest Cenozoic–Mesozoic evaporite basins, with preserved salt volumes greater than 10⁴ km³, are listed in Table 1; references are given in Supplementary Data. Continental reconstructions from this interval benefit from combined analysis of palaeomagnetic data from exposed rocks and seafloor anomalies (see, for example, ref. 24). The basins are distributed primarily through the 15–35° arid latitude belt as determined previously, with a 23 ± 4° (95% confidence) volume-weighted mean latitude (Fig. 1) that is independent of icehouse–greenhouse global climate state and the reversal frequency of geomagnetic polarity. Consistency of this result with modern climate zones bolsters the use of ancient evaporites as palaeolatitude proxies and supports a predominantly, if not entirely, GAD geomagnetic field model for the past 230 Myr (refs 14, 25). Detailed palaeomagnetic work on Late Triassic sedimentary rocks, spanning a range of palaeoclimatic zones in the north Atlantic region, also supports a GAD geomagnetic field model²⁶.

Palaeozoic evaporites are widespread and voluminous, as reviewed comprehensively by Zharkov²⁷. Although no seafloor-spreading data exist before the Jurassic period, Palaeozoic continental apparent polar wander (APW) paths are generally coherent enough to allow palaeolatitudes at specific ages to be calculated from running means or spline fits (see, for example, ref. 25). Palaeolatitude estimates vary

¹Department of Geology and Geophysics, 210 Whitney Avenue, Yale University, New Haven, Connecticut 06520-8109, USA.

Table 1 | Palaeomagnetic constraints on evaporite basin depositional latitudes

Evaporite basin (°N, °E)	Age (Myr)	Volume (km ³)	Palaeopole (°N, °E)	Dipole λ' (°)
Cenozoic–Mesozoic (0–250 Myr)				
1. Messinian (36, 018)	5	1,000,000	Present position	36 N
2. Red Sea (21, 038)	10	900,000	BC02 mean (85, 174)	18 ± 2 N
3. SW Iran (31, 047)	20	300,000	BC02 mean (84, 229)	25 ± 3 N
4. S Mozambique (–23, 035)	20	27,000	BC02 mean (84, 176)	28 ± 3 S
5. E China (34,115)	40	20,000	BC02 mean (81, 162)	40 ± 3 N
6. Rus, Arabia (25, 050)	50	200,000	BC02 mean (75, 237)	10 ± 3 N
7. N Sahara (32, 008)	90	32,000	BC02 mean (67, 249)	19 ± 5 N
8. Indochina (16, 104)	100	50,000	Y+01 mean (57, 170)	26 ± 4 N
9. S Atlantic (–3, 010)	120	35,000	BC02 mean (54, 261)	13 ± 3 S
10. Hith, Arabia (24, 050)	150	360,000	BC02 mean (47, 268)	11 ± 6 S
11. Central Asia (43, 070)	150	250,000	BC02 mean (75, 160)	41 ± 7 N
12. Andes (35, 290)	160	40,000	BC02 mean (89, 264)	35 ± 5 S
13. Gulf of Mexico (27, 265)	160	2,400,000	BC02 mean (74, 150)	19 ± 5 N
14. Alan, Arabia (33, 044)	180	20,000	BC02 mean (61, 270)	11 ± 5 N
15. Tanzania (–10, 040)	200	150,000	BC02 mean (62, 252)	33 ± 4 S
16. N Sahara (35, 355)	200	710,000	BC02 mean (62, 252)	25 ± 4 N
17. Keuper (52, 005)	225	50,000	T+01 mean (56, 132)	26 ± 3 N
18. Jilh, Arabia (29, 049)	230	120,000	TV02 mean (49, 254)	9 ± 11 S
19. S China (29, 106)	230	80,000	Z+96 mean (46, 218)	6 ± 7 N
Cenozoic–Mesozoic volume-weighted mean (95% confidence)				23 ± 4
Permian–Carboniferous (250–360 Myr)				
1. Zechstein (54, 011)	250	200,000	T+01 mean (52, 155)	20 ± 2 N
2. Khuff, Arabia (27, 050)	260	75,000	TC04,15% G3 (55, 158)	23
			TV02 mean (43, 257)	16 ± 11 S
3. E European (52, 052)	270	1,100,000	TC04, interpolated (36, 243)	26
			T+01 mean (45, 165)	23 ± 3 N
4. Peru–Bolivia (–13, 291)	270	62,000	TC04,12.5% G3 (45, 167)	22
			TV02 mean (68, 174)	22 ± 9 S
5. Midcontinental USA (41, 257)	270	81,000	TC04,12.5% G3 (62, 161)	30
			T+01 mean (44, 126)	6 ± 3 N
6. Amazon (–3, 310)	300	25,000	TC04, 12.5% G3 (44, 128)	7
			TV02 mean (60, 174)	24 ± 6 S
7. Sverdrup (79, 264)	315	120,000	TC04, 10% G3 (55, 158)	33
			V93 mean (28, 129)	20 ± 4 N
8. Canadian Maritime (46, 298)	340	46,000	TC04, 10% G3 (36, 127)	28
			TC04 mean (19, 118)	25 S
Permian–Carboniferous volume-weighted mean (95% confidence)				34
Recalculated for G3 contributions (95%)				21 ± 4
				23 ± 4
Devonian–late Ediacaran (360–600 Myr)				
1. E European (57, 040)	370	1,100,000	V93 mean (27, 149)	13 ± 8 N
2. Taimyr (74, 115)	370	18,000	TC04 mean (10, 117)	26 N
3. W Canada (55, 250)	390	86,000	V93 mean (23, 110)	5 ± 8 S
4. Morsovo (53, 033)	400	81,000	CT02 mean (2, 145)	11 S
5. Michigan (41, 279)	420	29,000	CT02 mean (14, 125)	30 S
6. Canning (–21, 124)	440	26,000	CT02 mean (19, 201)	5 N
7. Canadian Arctic (76, 265)	460	19,000	CT02 mean (13, 149)	7 N
8. Mackenzie (65, 234)	500	110,000	CT02 mean (4, 165)	12 N
9. Morocco–Iberia (34, 357)	520	50,000	Debated	—
10. Siberia (60, 105)	520	800,000	Tommotian–Toyonian	(8–25) S
11. Persian Gulf (22, 058)	545	500,000	Sinyai (–33, 326)*	13 ± 4 N
12. Salt Range (33, 073)	~550	240,000	Khewra (–28, 032)	18 ± 11 S
Devonian–Ediacaran volume-weighted mean (95% confidence)				14 ± 2
Precambrian (pre-Ediacaran; >600 Myr)				
1a. Skillogee (–31, 138)	~770	25,000	Mundine Well (46, 135)*	13 ± 4
1b. Curdimurka (–30, 138)	~785	50,000	None	–
2a. Kilian–Redstone River (68, 238)	~770	30,000	Direct (22, 151)	21 ± 9
2b. Minto Inlet (68, 238)	~800	90,000	L. Dal Basinal (–16, 141)	17 ± 3
3. Duruchaus (–22, 018)	~800	15,000	None	–
4. Copperbelt (–12, 027)	~830?	25,000	None	–
5. Centralian (–25, 129)	~830	140,000	Browne (–44, 312)	20 ± 5
6. Borden (73, 278)	~1,200	15,000	Strathcona S. (8, 204)	12 ± 3
7. Char/Douik (23, 352)	~1,200?	8,000	None	–
8. Belt (48, 245)	1,460	10,000	Belt B1-4 (–25, 214)*	12 ± 2
9. Discovery (–25, 118)	~1,500	≤2,800	None	–
10a. Balbirini (–17, 136)	1,610	2,500	Direct (–66, 178)	34 ± 6
10b. Lynott (–17, 136)	1,635	3,000	Direct (–75, 163)	30 ± 6
10c. Myrtle (–19, 138)	1,645	13,000	Emmerugga (–79, 203)	23 ± 6
10d. Mallapunyah (–19, 138)	1,660	5,000	Direct (–35, 214)	22 ± 3
10e. Corella (–21, 140)	1,740	2,000	Peters Creek (–26, 221)*	17 ± 5
11. Stark (62, 248)	~1,870	30,000	Stark (–15, 212)	8 ± 8
12. Rocknest (67, 246)	~1,950	1,000	Western R. (14, 341)	11 ± 9
13. Juderina (–26, 120)	~2,100	1,000	None	–
14. Tulomozero (63, 35)	~2,100	1,000	Kuetsyarvi (25, 301)*	20 ± 13
15. Chocolay (47, 275)	~2,250	4,500	Lorrain (–46, 268)	3 ± 3 (?)
Pre-Ediacaran volume-weighted mean (95% confidence)				17 ± 3

Evaporite localities represent the centres of the basins. Dipole λ' is the palaeolatitude of the basin centre assuming a GAD magnetic field; 95% uncertainties are from (A_{95}) or (dp) confidence limits on the palaeomagnetic poles. Poles are rotated into the coordinate reference frame of the evaporite basin, with Euler parameters from the cited sources, plus the Sinyai pole to Arabia from ref. 29. In some instances, vertical-axis rotations of the palaeomagnetic sample localities are known or suspected, but they are not corrected here. Abbreviations: G3, same-sign geocentric axial octupole component; BC02, ref. 24; CT02, ref. 47 (no uncertainties on means provided); T+01, ref. 48; TC04, ref. 39 (no uncertainties provided); TV02, ref. 25; V93, ref. 28; Y+01, ref. 49; Z+96, ref. 50. For further information and references on evaporite deposits and palaeomagnetic poles, see Supplementary Information.

* Individual palaeomagnetic studies derived wholly or partly from volcanic rocks.

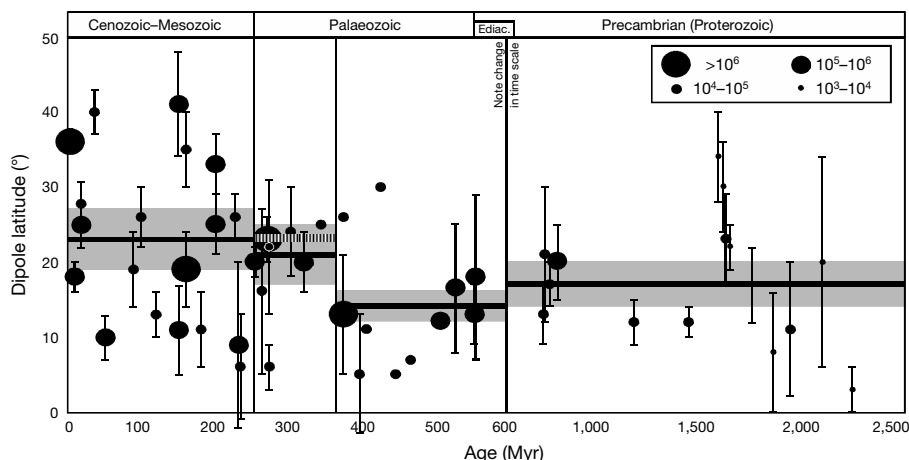


Figure 1 | Evaporite palaeomagnetic latitudes through time, assuming a GAD field as a null hypothesis. Individual basins are schematically scaled to evaporite volume (circle sizes are scaled in km^3 as in the inset) and shown with 95% uncertainty limits in palaeolatitude as listed in Table 1. Volume-weighted means (solid bars) and 95% error envelopes (shaded) are divided into the four age intervals as discussed in text. A corrected mean for best-fit octupole components can be calculated for the late Palaeozoic (dashed bar; uncertainty excluded for clarity).

little with averaging method or data selection, except for the age intervals of the Silurian–Devonian and the Early Cambrian^{28,29}, at times of widespread evaporite deposition (Table 1). For basins of these ages, special care was taken to consult primary constituent palaeomagnetic studies in order to determine which latitude estimates are most reliable. In the present compilation, the Palaeozoic (and late Ediacaran) basins are separated into two groups: Permian–Carboniferous, for which quantitative estimates of octupolar field components are available through palaeogeographic consistency tests across large continents^{25,30}, and Devonian–Ediacaran, for which no such estimates currently exist. Eight Permian–Carboniferous deposits with volumes greater than 10^4 km^3 have a weighted-mean dipole palaeolatitude of $21 \pm 4^\circ$, nearly identical to the Cenozoic–Mesozoic mean and nearly unaffected when previously estimated octupolar components are included in the pole calculations (Table 1). Among 11 out of 12 large Devonian–Ediacaran evaporites with palaeomagnetic constraints, however, the weighted-mean dipole palaeolatitude of $14 \pm 2^\circ$ is substantially lower (Fig. 1) and is again independent of global climate state and geomagnetic reversal frequency. As discussed below, this result is anomalous and may indicate exceptional periods of non-uniformitarian geomagnetism and/or palaeoclimate.

An extensive literature survey of described Precambrian evaporite deposits highlights 21 examples of great thickness and basin-wide scale (Table 1), the largest of which have also appeared in previous compilations^{31,32}. Continuous APW paths are rarely determined for segments of Precambrian time, so the estimation of palaeolatitudes requires specific results from the evaporite units themselves, conformably bounded strata, or coeval rocks on the same palaeocontinent. Evaporites with volumes greater than 10^3 km^3 are restricted to the Proterozoic era, as old as about 2,250 Myr. The volume-weighted mean inclination from the 15 palaeomagnetically constrained basins among this group ($31.6 \pm 2.1^\circ$ (s.e.m.)) corresponds to a GAD palaeolatitude of $17 \pm 3^\circ$ (95% confidence). This result is slightly lower than the Cenozoic–Mesozoic baseline value, with possible causes discussed below.

Testing high obliquity

Three published general-circulation climate models have reported evaporation–precipitation trends as a function of latitude on a high-obliquity planet with otherwise Earth-like parameters^{33–35}. These simulations indicate a near-isothermal to weakly reversed mean-annual zonal climate pattern, depending on whether the oceans are allowed to store and release heat through the annual cycle. Whenever there is any annually averaged evaporative peak present, the models locate this maximum at a single arid belt centred on the Equator. In conjunction with an assumed GAD field, these simulations would predict a unimodal peak of evaporite palaeomagnetic latitudes at 0° , with a large standard deviation.

Precambrian evaporite latitudes on Earth show a significantly non-zero mean with low variance (Table 1), consistent with a ‘normal’ climate gradient on a low-obliquity world. A zero mean could be obtained if half of the palaeolatitudes were considered to represent the opposite hemisphere (geomagnetic polarity being generally unknown in the Precambrian), but even then the total distribution would be significantly bimodal, in contrast with the high-obliquity model predictions. For some basins the palaeomagnetic record is complete enough to show continental motion through alternate times of evaporite and carbonate deposition, in patterns consistent with latitude crossings between subtropical arid and equatorial humid zones (Fig. 2). The Precambrian volume-weighted mean evaporite basin latitude ($17 \pm 3^\circ$) is indistinguishable from that of the entire Phanerozoic eon ($20 \pm 3^\circ$). Given that Phanerozoic obliquity is universally considered to have modern-like values⁹, the palaeomagnetic record of evaporites thus argues for uniformitarian orbital dynamics of the Earth–Moon system at least as old as the Palaeoproterozoic era.

Testing non-dipole field components

The second application of evaporite palaeolatitudes considers the long-term structure of the geomagnetic field. Although one can conceive of an infinitely complex hypothetical Precambrian geomagnetic field topology, only axial components can be distinguished through broad palaeoclimatic comparisons spanning billions of years. Because palaeomagnetic polarity is generally unknown among sparse Precambrian results, tests for subsidiary geomagnetic field components are practically limited to antisymmetric harmonics such as the geocentric axial octupole. Figure 3 shows the effect on surface inclinations of 0–40% subsidiary zonal octupole of the same sign relative to an otherwise pure GAD field³⁶, for the 15° and 35° latitudinal bounds of the modern arid zones and for the mean depositional latitudes of Cenozoic–Mesozoic evaporite basins. Only the Devonian–Ediacaran subset shows a significantly shallower mean inclination at the 99% confidence level (Student’s *t*-test), which could be interpreted as a persistent octupolar component of 19% relative to the dipole. This value is broadly similar in magnitude to that of the aggregate database-wide inclination tests of Kent and Smethurst¹⁴, who found a best-fit same-sign octupole component of about 25% for pre-Mesozoic data. Using the Cenozoic–Mesozoic evaporite inclination baseline, an octupolar contribution of this magnitude shifts the expected inclination range to $21.9 \pm 2.5^\circ$ (s.e.m.). Although pre-Ediacaran data are slightly shallower than the Cenozoic–Mesozoic GAD baseline, they are significantly different at the 99% confidence level (Student’s *t*-test) from that 25% octupole-adjusted baseline value. This suggests that substantial non-GAD components were not present in the long-term Precambrian geomagnetic field, and it supports the alternative explanation for database-wide inclination shallowing proposed by Kent and Smethurst¹⁴.

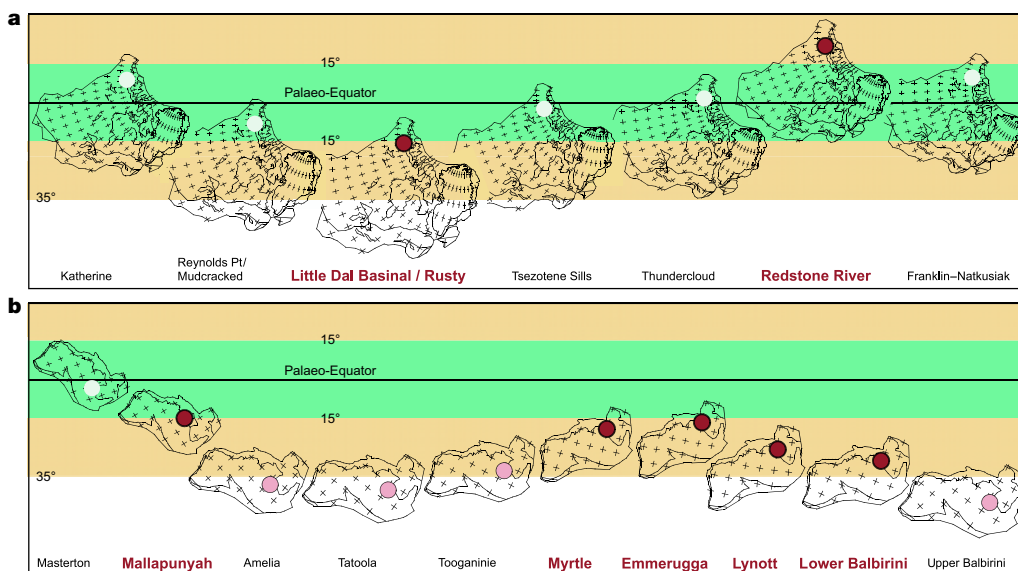


Figure 2 | Two sporadically evaporitic pre-Ediacaran basins that are palaeomagnetically well constrained throughout their depositional histories. The horizontal axis denotes stratigraphic development, with younger to the right. Latitude zones are colour-coded for the present-day humid tropics and arid subtropics, on a Mercator projection. Darker, lighter, and white circles indicate stratigraphic intervals of major, minor, and absent-to-negligible evaporites, respectively. **a**, Neoproterozoic Amundsen basin, Laurentia⁴⁵. **b**, Palaeoproterozoic McArthur basin, northern Australia⁴⁶.

namely that continental positions tend on average towards low latitudes by means of true polar wander through supercontinental cycles³⁷. Such a tendency would similarly impart a low-latitude bias to the preserved subset of rift-related evaporites, but the largest basins considered here are found in variable tectonic settings that should be independent of supercontinental episodicity.

A significant octupole component (about 10–15% relative to GAD) was determined by consistency tests on palaeomagnetic data from the late Palaeozoic continents Laurussia³⁰ and Gondwanaland²⁵, and this also resolved some long-standing overlap problems with reconstructions of Pangaea between those two plates (a pure GAD model requires a large Permian megashear between the two continents³⁸). The present compilation of evaporite inclinations does not provide further insight into this debate, because error-min-

imizing octupole field corrections for Permian to Late Carboniferous APW paths³⁹ have no significant effects on the volume-weighted mean for evaporites of that interval (Table 1). Octupole-corrected early Palaeozoic APW paths are not yet available for all continents, but such calculations will help to determine whether the early part of the Palaeozoic era was a time of departure from an otherwise uniformitarian GAD field, perhaps as a result of particular convective patterns at the core–mantle boundary¹⁵.

Other factors may have contributed to the substantially shallowed mean inclination for Devonian–Ediacaran evaporites, and the slightly shallowed mean for pre-Ediacaran evaporites: first, narrowing of the Hadley zones due to a faster rotation rate⁴⁰; second, a shallow bias of palaeomagnetic inclinations due to sedimentary magnetization processes⁴¹; and third, particular continental distributions conducive to aridity in deep tropical regions^{23,42}. The first process may have contributed to the shallowing of Precambrian inclinations but not substantially to the more marked early Palaeozoic anomaly. Regarding the second process, the pre-Ediacaran mean remains low even when only igneous-based palaeomagnetic results are considered (Table 1). The significantly shallowed early Palaeozoic mean evaporite palaeolatitude could be a result of the third process, because several of the larger basins reconstruct near the vertex of a proto-Tethys-like sea³⁹, which may have been influenced by monsoon-like circulation in the same manner that produces anomalously dry conditions in present-day eastern Africa²³. The slightly shallowed pre-Ediacaran mean inclination could also reflect one or more instances of similar palaeoclimatic peculiarities.

Implications

First-order consistency of Precambrian evaporite basin palaeomagnetic latitudes with modern arid zones and Cenozoic–Mesozoic evaporite palaeolatitudes confirms the GAD hypothesis for the Earth's magnetic field since about 2,000 Myr ago, about four times longer than known from previous palaeoclimatic–palaeomagnetic global comparisons^{2,21,22}. Because no voluminous evaporites are preserved from the times of Proterozoic ice ages, the present study allows the possibility of relatively short-term geomagnetic departures from the GAD model, as a potential contributor to the low glacial palaeolatitudes. However, those departures would need to be extreme to produce the near-equatorial results, and, on the contrary, strata-bound antiparallel palaeomagnetic polarity reversals from within some of the Proterozoic glaciogenic successions⁴³ seem to reflect 'typical' geomagnetic behaviour of more recent times. Short-term departures from low Earth obliquity are not possible because of the stabilizing influence of the Moon¹¹.

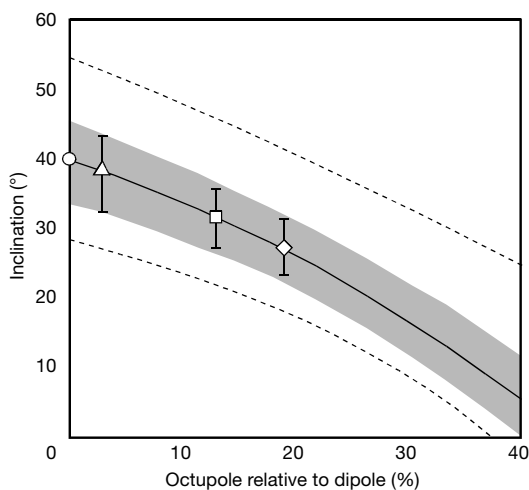


Figure 3 | Plot of magnetic inclination against same-sign geomagnetic octupole component relative to a purely GAD field. The values for the inclination biases were calculated from the equation in ref. 36. Dashed curves indicate the declining inclination values, according to increasing octupole contributions, that would correspond to 15° and 35° latitude limits of present arid subtropical zones. The solid curve corresponds to the same biasing effect on inclinations computed from the mean latitude of Cenozoic–Mesozoic evaporite basins and a pure GAD field (circle), with its 95% error envelope (shaded). Permian–Carboniferous (triangle), Devonian–Ediacaran (diamond), and pre-Ediacaran (square) mean inclinations and their 95% uncertainty limits, when aligned with the Cenozoic–Mesozoic baseline curve, provide quantifiable estimates of ancient geomagnetic octupole field components.

Earth's evaporite record extends in abundance only through the Proterozoic Eon^{31,32}; studies such as this are therefore unlikely to characterize the growth of Earth's geomagnetic field, which seems to have occurred before 2,450 Myr ago⁴⁴. Conversely, reliable palaeomagnetic data from Archaean sedimentary rocks are currently so sparse as to preclude comprehensive treatment of any palaeoclimatic indicator from that interval, evaporitic or otherwise. Nonetheless, the present compilation addresses several important issues of deep time: first, that the high-obliquity hypothesis has failed an important observational test that is independent of Earth's Precambrian glacial record; second, that near-equatorial palaeomagnetic latitudes of Precambrian glacial deposits are not likely to be substantially biased by non-dipolar field components, unless such components were of exceptional magnitude and short duration coincident with ice ages; and third, that—according to the available tests—Earth's time-averaged magnetic field may be approximated by a geocentric axial dipole for most of the past 2 Gyr. Success of the uniformitarian principle in explaining evaporite distributions implies that its failure, with regard to low-latitude Precambrian glaciation, is most probably caused by anomalies in palaeoclimatic rather than geophysical processes.

Received 4 April; accepted 4 September 2006.

- Evans, D. A. D. Stratigraphic, geochronological, and paleomagnetic constraints upon the Neoproterozoic climatic paradox. *Am. J. Sci.* **300**, 347–433 (2000).
- Evans, D. A. D. A fundamental Precambrian–Phanerozoic shift in Earth's glacial style? *Tectonophysics* **375**, 353–385 (2003).
- Hoffman, P. F., Kaufman, A. J., Halverson, G. P. & Schrag, D. P. A Neoproterozoic snowball Earth. *Science* **281**, 1342–1346 (1998).
- Hoffman, P. F. & Schrag, D. P. The snowball Earth hypothesis: Testing the limits of global change. *Terra Nova* **14**, 129–155 (2002).
- Kirschvink, J. L. in *The Proterozoic Biosphere: A Multidisciplinary Study* (eds Schopf, J. W. & Klein, C.) 51–52 (Cambridge Univ. Press, Cambridge, 1992).
- Hyde, W. T., Crowley, T. J., Baum, S. K. & Peltier, W. R. Neoproterozoic 'snowball Earth' simulations with a coupled climate/ice sheet model. *Nature* **405**, 425–429 (2000).
- Poulsen, C. J. Absence of a runaway ice-albedo feedback in the Neoproterozoic. *Geology* **31**, 473–476 (2003).
- Williams, G. E. Late Precambrian glacial climate and the Earth's obliquity. *Geol. Mag.* **112**, 441–465 (1975).
- Williams, G. E. History of the Earth's obliquity. *Earth Sci. Rev.* **34**, 1–45 (1993).
- Vanyo, J. P. & Awramik, S. M. Stromatolites and Earth–Sun–Moon dynamics. *Precamb. Res.* **29**, 121–142 (1985).
- Levrard, B. & Laskar, J. Climate friction and the Earth's obliquity. *Geophys. J. Int.* **154**, 970–990 (2003).
- Pais, M. A., Le Mouél, J. L., Lambeck, K. & Poirer, J. P. Late Precambrian paradoxical glaciation and obliquity of the Earth: A discussion of dynamical constraints. *Earth Planet. Sci. Lett.* **174**, 155–171 (1999).
- Williams, D. M., Kasting, J. F. & Frakes, L. A. Low-latitude glaciation and rapid changes in the Earth's obliquity explained by obliquity-oblateness feedback. *Nature* **396**, 453–455 (1998).
- Kent, D. V. & Smethurst, M. A. Shallow bias of paleomagnetic inclinations in the Paleozoic and Precambrian. *Earth Planet. Sci. Lett.* **160**, 391–402 (1998).
- Bloxham, J. Sensitivity of the geomagnetic axial dipole to thermal core–mantle interactions. *Nature* **405**, 63–65 (2000).
- Olson, P. & Christensen, U. R. The time-averaged magnetic field in numerical dynamos with non-uniform boundary heat flow. *Geophys. J. Int.* **151**, 809–823 (2002).
- Irving, E. Palaeomagnetic and paleoclimatological aspects of polar wandering. *Geophys. Pura Appl.* **33**, 23–41 (1956).
- Irving, E. & Briden, J. C. Palaeolatitude of evaporite deposits. *Nature* **196**, 425–428 (1962).
- Opdyke, N. D. in *Continental Drift* (ed. Runcorn, S. K.) 41–65 (Academic, New York, 1962).
- Borchert, H. & Muir, R. O. *Salt Deposits* (Van Nostrand, London, 1964).
- Drewry, G. E., Ramsay, A. T. S. & Smith, A. G. Climatically controlled sediments, the geomagnetic field, and trade wind belts in Phanerozoic time. *J. Geol.* **82**, 531–553 (1974).
- Gordon, W. A. Distribution by latitude of Phanerozoic evaporite deposits. *J. Geol.* **83**, 671–684 (1975).
- Parrish, J. T., Ziegler, A. M. & Scotese, C. R. Rainfall patterns and the distribution of coals and evaporites in the Mesozoic and Cenozoic. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **40**, 67–101 (1982).
- Besse, J. & Courtillot, V. Apparent and true polar wander and the geometry of the geomagnetic field over the last 200 Myr. *J. Geophys. Res.* **107**, doi:10.1029/2000JB000050 (2002).
- Torsvik, T. H. & Van der Voo, R. Refining Gondwana and Pangea palaeogeography: estimates of Phanerozoic non-dipole (octupole) fields. *Geophys. J. Int.* **151**, 771–794 (2002).
- Kent, D. V. & Tauxe, L. Corrected Late Triassic latitudes for continents adjacent to the North Atlantic. *Science* **307**, 240–244 (2005).
- Zharkov, M. A. *History of Paleozoic Salt Accumulation* (Springer, Berlin, 1981).
- Van der Voo, R. *Paleomagnetism of the Atlantic, Tethys and Iapetus Oceans* (Cambridge Univ. Press, Cambridge, 1993).
- McElhinny, M. W., Powell, C. M. & Pisarevsky, S. A. Paleozoic terranes of eastern Australia and the drift history of Gondwana. *Tectonophysics* **362**, 41–65 (2003).
- Van der Voo, R. & Torsvik, T. H. Evidence for late Paleozoic and Mesozoic non-dipole fields provides an explanation for the Pangea reconstruction problems. *Earth Planet. Sci. Lett.* **187**, 71–81 (2001).
- Pope, M. C. & Grotzinger, J. P. Paleoproterozoic Stark Formation, Athapuscow basin, northwest Canada: Record of cratonic-scale salinity crisis. *J. Sedim. Res.* **73**, 280–295 (2003).
- Warren, J. K. *Evaporites: Their Evolution and Economics* (Blackwell Science, Oxford, 1999).
- Hunt, B. G. The impact of large variations of the Earth's obliquity on the climate. *J. Meteorol. Soc. Jpn* **60**, 309–318 (1982).
- Jenkins, G. S. Global climate model high-obliquity solutions to the ancient climate puzzles of the Faint Young Sun Paradox and low-altitude Proterozoic Glaciation. *J. Geophys. Res.* **105**, 7357–7370 (2000).
- Jenkins, G. S. High-obliquity simulations for the Archean Earth: Implications for climatic conditions on early Mars. *J. Geophys. Res.* **106**, 32903–32913 (2001).
- Livermore, R. A., Vine, F. J. & Smith, A. G. Plate motions and the geomagnetic field. I. Quaternary and late Tertiary. *Geophys. J. R. Astron. Soc.* **73**, 153–171 (1983).
- Evans, D. A. D. True polar wander and supercontinents. *Tectonophysics* **362**, 303–320 (2003).
- Muttoni, G. et al. Early Permian Pangea 'B' to Late Permian Pangea 'A'. *Earth Planet. Sci. Lett.* **215**, 379–394 (2003).
- Torsvik, T. H. & Cocks, L. R. M. Earth geography from 400 to 250 Ma: a palaeomagnetic, faunal and facies review. *J. Geol. Soc. Lond.* **161**, 555–572 (2004).
- Hunt, B. G. The effects of past variations of the Earth's rotation rate on climate. *Nature* **281**, 188–191 (1979).
- Tauxe, L. Inclination flattening and the geocentric axial dipole hypothesis. *Earth Planet. Sci. Lett.* **233**, 247–261 (2005).
- Hallam, A. Continental humid and arid zones during the Jurassic and Cretaceous. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **47**, 195–223 (1984).
- Raub, T. D. & Evans, D. A. D. Magnetic reversals in basal Ediacaran cap carbonates: A critical review. *Eos* **87** (36), abstract G41B-02 (2006).
- Smirnov, A. V., Tarduno, J. A. & Pisakin, B. N. Paleointensity of the early geodynamo (2.45 Ga) as recorded in Karelia: A single-crystal approach. *Geology* **31**, 415–418 (2003).
- Park, J. K. & Jefferson, C. W. Magnetic and tectonic history of the Late Proterozoic Upper Little Dal and Coates Lake Groups of northwestern Canada. *Precamb. Res.* **52**, 1–35 (1991).
- Idnurm, M. Towards a high resolution Late Palaeoproterozoic – earliest Mesoproterozoic apparent polar wander path for northern Australia. *Aust. J. Earth Sci.* **47**, 405–429 (2000).
- Cocks, L. R. M. & Torsvik, T. H. Earth geography from 500 to 400 million years ago: a faunal and paleomagnetic review. *J. Geol. Soc. Lond.* **159**, 631–644 (2002).
- Torsvik, T. H., Van der Voo, R., Meert, J. G., Mosar, J. & Walderhaug, H. J. Reconstructions of the continents around the North Atlantic at about the 60th parallel. *Earth Planet. Sci. Lett.* **187**, 55–69 (2001).
- Yang, Z., Yin, J., Sun, Z., Otofujii, Y. & Sato, K. Discrepant Cretaceous paleomagnetic poles between Eastern China and Indochina: a consequence of the extrusion of Indochina. *Tectonophysics* **334**, 101–113 (2001).
- Zhao, X., Coe, R. S., Gilder, S. A. & Frost, G. M. Palaeomagnetic constraints on the palaeogeography of China: implications for Gondwanaland. *Aust. J. Earth Sci.* **43**, 643–672 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements I thank J. Emerson, K. Grey, J. Grotzinger, P. Hoffman, D. Kent, R. Rainbird, Tim and Theresa Raub, S. Sherwood and P. Southgate for discussions, and R. Van der Voo for constructive comments on the manuscript. The David and Lucile Packard Foundation provided support.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Correspondence and requests for materials should be addressed to D.A.D.E. (dai.evans@yale.edu).

ARTICLES

Long-term motor cortex plasticity induced by an electronic neural implant

Andrew Jackson¹, Jaideep Mavoori² & Eberhard E. Fetz¹

It has been proposed that the efficacy of neuronal connections is strengthened when there is a persistent causal relationship between presynaptic and postsynaptic activity. Such activity-dependent plasticity may underlie the reorganization of cortical representations during learning, although direct *in vivo* evidence is lacking. Here we show that stable reorganization of motor output can be induced by an artificial connection between two sites in the motor cortex of freely behaving primates. An autonomously operating electronic implant used action potentials recorded on one electrode to trigger electrical stimuli delivered at another location. Over one or more days of continuous operation, the output evoked from the recording site shifted to resemble the output from the corresponding stimulation site, in a manner consistent with the potentiation of synaptic connections between the artificially synchronized populations of neurons. Changes persisted in some cases for more than one week, whereas the output from sites not incorporated in the connection was unaffected. This method for inducing functional reorganization *in vivo* by using physiologically derived stimulus trains may have practical application in neurorehabilitation after injury.

Learning is associated with a long-term reorganization of sensory and motor representations in the neocortex^{1–3}. Acquisition of motor skills can alter the somatotopic map of limb movements in sensorimotor areas^{4,5}, and plastic changes have been implicated in the recovery from disorders such as stroke⁶ and incomplete spinal cord injury⁷. Hebb⁸ postulated that synaptic efficacy between two neurons is strengthened if the first repeatedly contributes to firing the second; since then, long-term potentiation of motor cortical synapses^{9–11} and associated expansion of movement representations^{12,13} have been induced with tetanic stimulus trains delivered to individual sites. Long-term potentiation is widely assumed to reflect a learning mechanism, but its physiological relevance and relationship to hebbian plasticity are often inferred indirectly¹⁴. More recently, spike timing-dependent plasticity consistent with Hebb's criterion of a causal relationship between presynaptic and postsynaptic firing has been

described at the cellular level^{15,16}. However, it remains to be shown that such a mechanism can cause stable, functional reorganization of cortical maps *in vivo* under normal behavioural conditions.

We are developing implantable electronic circuits¹⁷ (or Neurochips; Supplementary Fig. 1) for neural recording and stimulation in freely behaving animals that could provide prosthetic connections to replace or augment damaged pathways in the nervous system¹⁸. A Neurochip creates an artificial connection between two sites by using action potentials recorded on one electrode to trigger electrical stimuli delivered to another (Fig. 1a). Once configured with appropriate recording and stimulation parameters, the Neurochip operates autonomously, allowing connections to function continuously over days of unrestrained behaviour. Because the Neurochip creates a causal relationship between neural activities at connected sites, its long-term operation could also induce changes mediated by hebbian

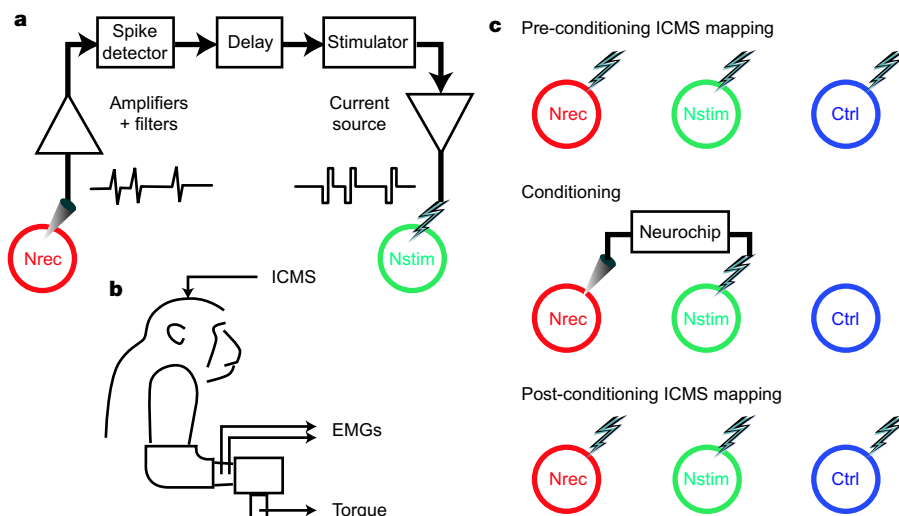


Figure 1 | Conditioning protocol and experimental design. **a**, Diagram of the artificial connection. Action potentials detected in the signal recorded from the Nrec electrode triggered electrical stimuli delivered to the Nstim electrode after a predefined delay. **b**, Experimental setup for testing output effects of ICMS on the right wrist. **c**, Experimental sequence of ICMS testing and conditioning with the Neurochip.

¹Department of Physiology and Biophysics and Washington National Primate Research Center, ²Department of Electrical Engineering, University of Washington, Seattle, Washington 98195, USA.

mechanisms. Here we describe stable reorganization of movement representations in the wrist area of the primary motor cortex (M1) in monkeys resulting from artificial connections between pairs of electrodes in a chronically implanted array. We found that the motor output elicited from recording sites shifted towards the output evoked from stimulation sites. These changes occurred only when stimuli were delivered within 50 ms of recorded spikes. The output evoked from neighbouring control electrodes was unchanged. This shows that natural patterns of cortical spiking *in vivo* during normal behaviour can lead to input-specific hebbian plasticity when paired with appropriate stimulation. Plastic changes arising from these artificial connections could have clinical applications in rehabilitation after motor injury.

Conditioning with an artificial connection

The output effects in the contralateral wrist evoked from cortical electrodes before, during and after conditioning with the artificial connection were tested with daily intracortical microstimulation (ICMS, Fig. 1b, c). Figure 2a shows the pre-conditioning average trajectories of isometric wrist torque (dashed lines) for 200 ms after a train of stimuli delivered separately to each of three electrodes, designated Neurochip recording (Nrec), Neurochip stimulation (Nstim) and control (Ctrl). The mean torque, indicated by solid arrows, was towards the flexion direction for Nrec and Ctrl, and in the radial-extension direction for Nstim. Figure 2b shows average rectified electromyogram (EMG) responses to ICMS in three wrist muscles: extensor carpi radialis (ECR), flexor carpi radialis (FCR) and flexor carpi ulnaris (FCU).

The Neurochip was then programmed to deliver a single stimulus pulse to Nstim 5 ms after every action potential detected at Nrec (Supplementary Fig. 2a). The cell at this site fired preferentially with flexion during torque-tracking and its activity was correlated with both wrist flexor and extensor muscles during free behaviour (Supplementary Fig. 2f, g). The stimulus intensity (40 μ A) was below the movement threshold for a single pulse but was sufficient for occasional bursts of cell activity to elicit small muscle twitches while the monkey sat at rest. This connection produced no noticeable disruption of the control of active movements, probably because the effect of stimulation was weak in comparison with neural activity generating volitional movement. The artificial connection operated continuously for two days while the monkey moved unrestrained about the home cage. During this period, the Neurochip also recorded the stimulation rate over consecutive 1-s bins (Supplementary Fig. 2b). The mean rate of stimulation was 19 Hz during daytime behaviour and 9 Hz during the night. This pattern was consistent with our observations of robust correlations between the firing rate of M1 neurons and muscle activity during natural behaviour, and cycles of activity and quiescence during sleep¹⁹.

After two days of conditioning, the mean torque generated by ICMS at Nrec had shifted to the radial direction, towards the output effect produced from Nstim (Fig. 2c). The torque now followed a curved trajectory aligned initially with the Nstim trajectory, before returning to the flexion direction. This curved path was produced by a new pattern of muscle responses (Fig. 2d), which included initial activation of the extensor muscle ECR (black arrows), previously elicited only by ICMS at Nstim. The output from the control electrode was unaffected, indicating that the plastic change was caused by the pairing of neural activity at the Nrec site with the stimulation of Nstim. In this case, the response from the Nstim site had also increased slightly (Supplementary Fig. 3), but this was not generally so (see Supplementary Information). Figure 2e plots the angle of mean torque produced by ICMS at each site for the days before, during and after conditioning. The changes at the Nrec site developed gradually over the two days of conditioning and subsequently remained stable for one week.

Summary of conditioning sessions

We investigated the effect of creating artificial connections between 17 different pairs of electrodes over separate conditioning sessions in two monkeys (Y, eight sessions; K, nine sessions). Conditioning lasted for 1–4 days (mean 1.6 days) with Nstim currents in the range 25–80 μ A (mean 48 μ A) and delays of 0, 1 and 5 ms interposed between spike and stimulus (Supplementary Table 1). ICMS effects were quantified by the angle of mean wrist torque relative to the pre-conditioning direction of Nstim effect. Control electrodes were chosen to have similar ICMS effects to those of the Nrec pre-conditioning. Figure 3a summarizes the direction of mean torque produced by ICMS before and after each session. Points lying on the dashed line of identity represent ICMS effects that were unaffected by conditioning; points lying below the line represent effects that moved towards the output effect of Nstim. Conditioning had no significant effect on the separation of Ctrl site effects from the Nstim direction, which changed by $2.1^\circ \pm 2.3^\circ$ (mean $\pm$ s.e.m., two-tailed

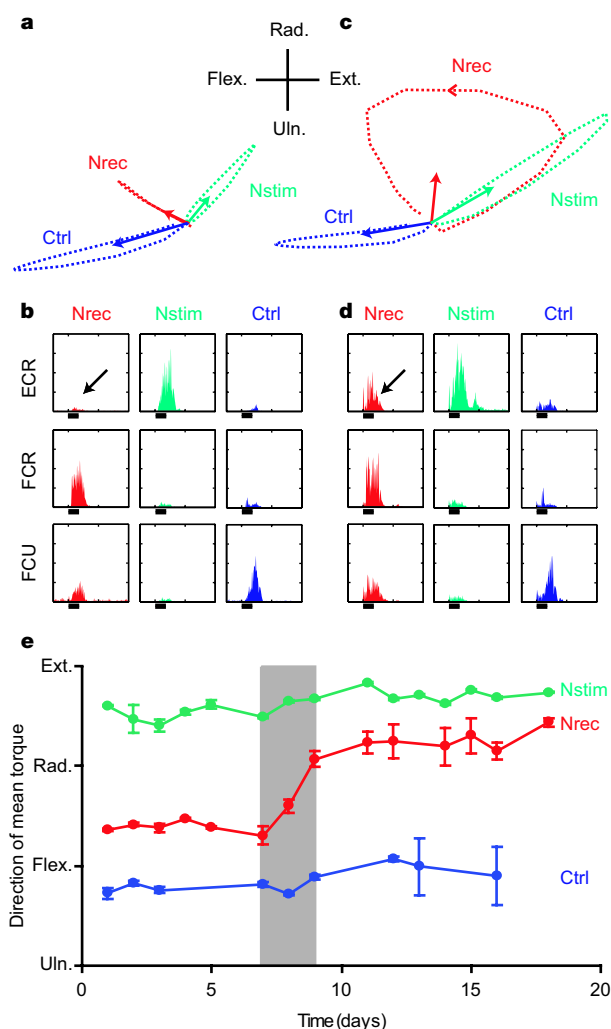


Figure 2 | Reorganization of motor output after conditioning. **a**, Average wrist torque responses to ICMS before conditioning. Arrows show the means of each 200-ms trajectory (dashed lines). Calibration bars are ± 0.02 N m (Nrec and Ctrl) and ± 0.08 N m (Nstim). **b**, Peristimulus averages of the rectified EMG response. The full axis lengths are 250 ms (x) and 0.4 mV (y). The bar indicates the ICMS train. **c**, **d**, Data after two days of conditioning with an artificial connection. Arrows indicate increased ECR response from Nrec after conditioning. **e**, Angle of torque response over 18 days. Shading indicates the conditioning period. Error bars show s.e.m. ICMS, 13 pulses at 300 Hz; current, 30 μ A (Nrec), 40 μ A (Nstim) and 50 μ A (Ctrl). The data are given in session 3 of Supplementary Table 1 and are averages of 20 stimulus trains.

paired *t*-test, $P = 0.4$). In contrast, Nrec effects moved towards the Nstim direction by $38^\circ \pm 9^\circ$ (mean $\pm$ s.e.m., $P = 0.0005$). In 13 of 17 individual sessions, Nrec effects rotated by an angle greater than the 95th centile of the Ctrl distribution (15°). The angular shift per day of conditioning was comparable for both animals (Y, $24^\circ \pm 8^\circ$; K, $25^\circ \pm 7^\circ$; mean $\pm$ s.e.m.). Figure 3b shows average angular separations of Nrec and Ctrl effects from the Nstim direction before, just after and one day after the end of the conditioning period. Supplementary Fig. 4 shows example torque trajectories from these experiments. Supplementary Fig. 5 shows the one session in which the change at the Nstim site gradually wore off, six to eight days after the end of conditioning.

Dependence on spike–stimulus delay

The absence of conditioning effects at control sites suggests that the timing of stimulation relative to cell activity had a crucial function in inducing plasticity. Neurons distributed widely throughout the motor cortex exhibit correlated firing on the timescale of movements (typically several hundred milliseconds^{19,20}; Supplementary Figs 2g and 6a). The site specificity of our results therefore indicates that a more precise coincidence between spikes and stimulation was required to induce plastic changes. We tested this hypothesis in a further series of experiments with monkey K by introducing longer delays of 20–2,000 ms between the spike and stimulation pulse (see Supplementary Information). Figure 4 summarizes these data by

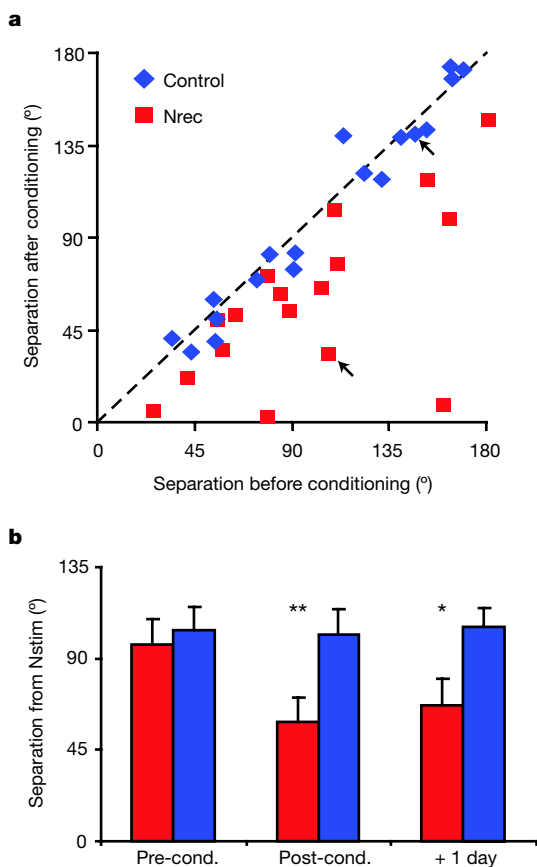


Figure 3 | Summary of conditioning results. **a**, Angular separation of Nrec and Ctrl ICMS effects from the pre-conditioning Nstim direction before and after conditioning for 17 sessions. Control data (blue) cluster around the dashed line of identity. Nrec data (red) fall below the line, indicating that the mean torque shifted towards the response elicited from Nstim. Arrows indicate data in Fig. 2. **b**, Average angular separation of Nrec (red bars) and Ctrl (blue bars) effects from the Nstim direction before conditioning. Error bars show s.e.m. '+1 day' describes one day after the end of conditioning. Asterisk, $P = 0.001$; two asterisks, $P = 0.0005$ ($n = 17$, two-tailed paired *t*-test relative to pre-conditioning data).

plotting the angular shift of the Nrec effect towards the Nstim direction per day of conditioning as a function of stimulus delay. Significant shifts ($P < 0.05$, two-tailed *t*-test) were obtained for intervals up to 50 ms, indicating that a coincidence of spike and stimulus within this window was required for inducing plasticity. The average angular shift for delays of 20 and 50 ms ($35^\circ \pm 5^\circ$) was slightly greater than had been obtained with the shorter delays ($24^\circ \pm 5^\circ$), but this difference was not significant ($P = 0.25$, two-tailed unpaired *t*-test).

Discussion

These results may be explained by the potentiation of horizontal pathways within the motor cortex¹⁰ such that ICMS delivered, after conditioning, to Nrec activates additional muscle groups through Nstim (Fig. 5). Alternatively, plasticity could occur at other cortical or subcortical targets of converging projections from Nrec and Nstim sites. Repetitive high-frequency stimulation has been shown to expand movement representations in the motor cortex of rats^{12,13}, but a general expansion of local Nstim effects cannot account for the changes we saw at Nrec sites for several reasons. First and foremost, outputs from neighbouring control electrodes were unaffected by conditioning. There was no significant difference in either the distance from the Nstim site (Nrec, 1.09 ± 0.16 mm; Ctrl, 0.95 ± 0.12 mm; mean $\pm$ s.e.m.; two-tailed unpaired *t*-test, $P = 0.5$) or ICMS currents used (Nrec, 68 ± 10 μ A, Ctrl, 60 ± 7 μ A, $P = 0.5$). Second, the angular shift of ICMS effects evoked from Nrec was not correlated with distance from the Nstim site (Pearson's $r = -0.0007$). Last, no shifts occurred when stimuli were delayed by more than 50 ms relative to the Nrec spikes, although the temporal pattern of stimulation was equivalent. These observations all indicate that the relative timing of Nrec spikes and Nstim stimulation was the critical factor for inducing plasticity.

Cellular studies of spike timing-dependent plasticity have shown that plasticity at individual synapses is triggered by Ca^{2+} influx, requiring both presynaptic glutamate release and postsynaptic depolarization to release the Mg^{2+} block of *N*-methyl-D-aspartate channels¹⁶. We propose that in our experiments the depolarization of local or downstream neurons by stimulation at Nstim induced the potentiation of synapses concurrently activated by spikes arriving from the Nrec site. Inputs from other sites (namely Ctrl electrodes) were not potentiated because the timing of this presynaptic activity had no consistent correlation with postsynaptic depolarization. Previous studies have shown that synaptic inputs activated after postsynaptic depolarization can be depressed^{15,16}, so although some fraction of spikes from control sites would have arrived during the window for synaptic potentiation, these could be cancelled by a comparable

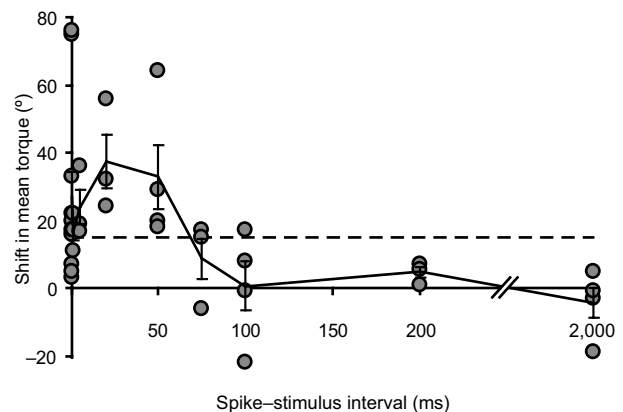


Figure 4 | Dependence of conditioning effects on delay between spikes and stimuli. The graph shows angular shift of Nrec effects towards Nstim effects per day of conditioning for different spike–stimulus intervals. The solid line connects the group means for each interval. Error bars show s.e.m. The dashed line indicates the 95th centile for control electrodes obtained from the previous experiment.

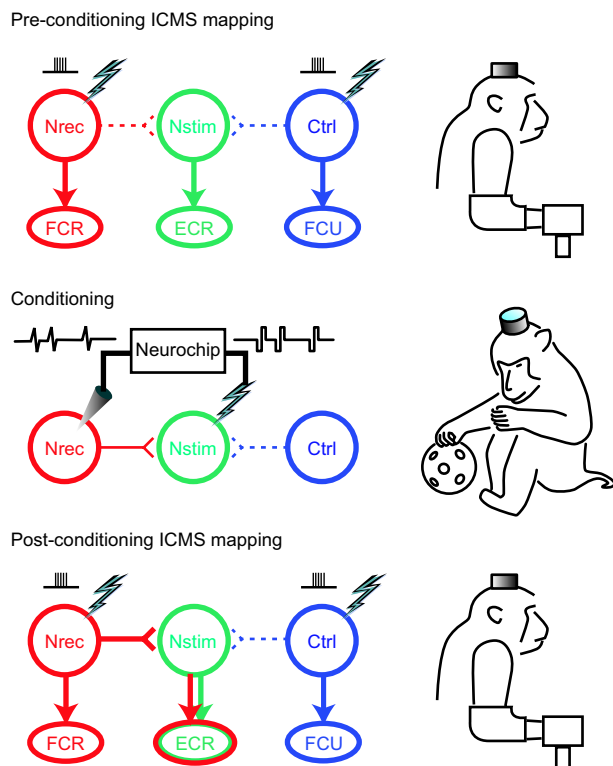


Figure 5 | Suggested mechanism for the conditioning effect documented in Fig. 2. Pre-conditioning ICMS predominantly activates distinct descending projections from Nrec to FCR, from Nstim to ECR, and from Ctrl to FCU. Conditioning during unrestrained behaviour induces a strengthening of horizontal connections between Nrec and Nstim. Post-conditioning ICMS now activates ECR by means of horizontal projections to Nstim, as well as activating FCR by means of the direct projection.

number of inputs arriving during the window for synaptic depression. This would also explain why delays of about 20 ms tended to produce the strongest conditioning effects. With shorter delays more Nrec spikes would have arrived after stimulation, during the window for synaptic depression.

Associative plasticity has been demonstrated previously with the use of paired stimulation of two input pathways in the cerebellum²¹, hippocampus²² and motor cortex²³. In contrast with those studies, we induced a functional reorganization by using *in vivo* spike activity at one site to trigger the stimulation of a second site. This constitutes a relatively direct test of Hebb's postulate, showing that natural patterns of neuronal firing can lead to input-specific plasticity when paired with appropriate postsynaptic depolarization during normal behaviour. It seems unlikely that the magnitude of this reorganization can be accounted for by altered projections only from the recorded neuron, because ICMS delivered to Nrec presumably activated populations of cells by means of local circuitry and temporal summation²⁴. However, neighbouring M1 neurons with similar output projections exhibit the maximal degree of synchronous discharge^{25–27} (Supplementary Fig. 6b), so during conditioning many spikes from this population will be temporally correlated within the coincidence window for synaptic potentiation.

Our method for inducing plasticity shares similarities with cellular conditioning protocols *in vivo* in sensory areas that pair spontaneous or evoked neuronal activity with appropriate sensory stimulation^{28–30}. However, the changes seen at the cellular level in those studies typically lasted for a few hours at most and were reversed by normal activity. The continuous conditioning over long periods of natural behaviour implemented by our Neurochip system may account for the strength and stability of the effects described here; they are consistent with the finding that multiple sessions are

required to induce stable long-term potentiation in rats *in vivo*¹¹. Furthermore, conditioning was associated with volitional movements rather than externally imposed activation and was continued during natural sleep, including rapid eye movement phases when motor cortical neurons can be highly active¹⁹. Sleep has been implicated in the consolidation of motor memory³¹, but the relative contributions of waking and sleeping periods to our results remain to be determined.

Artificial connections could provide a neural prosthesis to replace damaged pathways in the nervous system after injury¹⁸. Our results suggest that an additional rehabilitative consequence in cases of partial injury may be the strengthening of surviving projections between sites connected by the prosthesis. Functional reorganization is thought to be important in recovery from numerous movement disorders^{6,7}, and new stimulation protocols are being developed to aid this process^{32–34}. Stimulation in real time triggered from neural recordings during volitional movements could provide an effective method of selectively strengthening specific neural pathways during rehabilitation.

METHODS

See Supplementary Information for additional methods.

Subjects. Experiments were performed with two male *Macaca nemestrina* monkeys: Y (3 years old; weight 4.3 kg) and K (3 years old; weight 4.6 kg). All procedures were approved by the University of Washington Institutional Animal Care and Use Committee.

Neurochip implant. A full description of the Neurochip Brain–Computer Interface has been published previously¹⁷. The battery-powered circuit allows continuous long-term recording and stimulation during unrestrained behaviour through an array of 12 chronically implanted moveable tungsten microwire electrodes in M1 (diameter 50 μ m; impedance 0.5 M Ω ; interelectrode spacing 500 μ m). A microprocessor identified isolated action potentials from Nrec and instructed a stimulator circuit to deliver biphasic, constant-current pulses (0.2 ms per phase) to Nstim after a specified delay. Short sections of raw recording (sampled at 11.7 kHz) and stimulation rate in 1-s bins over the duration of conditioning were stored to on-board memory.

ICMS protocol. ICMS effects were documented with a current that was just above the threshold for eliciting a torque response before conditioning, and the same current was used throughout. The monkey sat with elbow and hand immobilized by padded restraints. A force transducer measured the two-dimensional isometric torque produced at the wrist in the flexion–extension and radial–ulnar directions. During some sessions, the EMG was recorded with pairs of stainless steel wires inserted transcutaneously into the wrist muscles. Torque and EMG were recorded at 5 kHz. Offline the torque trace was smoothed and downsampled to 100 Hz. Trains of 13 biphasic ICMS pulses (0.2 ms per phase) at 300 Hz were delivered at 2-s intervals. Peristimulus averages of torque and rectified EMG profiles were compiled from 100 ms before each stimulus to 500 ms after it. Traces in which the torque level before the stimulus exceeded 0.02 N m in any direction were excluded from the average. The trajectories in Fig. 2 connect the vector average of two-dimensional torque across stimulus trains for consecutive sample points up to 200 ms after stimulation. The vector average of this trajectory was used to determine the direction and magnitude of the mean torque response to ICMS at each site.

Received 1 May 2006; accepted 29 August 2006.

Published online 22 October 2006.

- Buonomano, D. V. & Merzenich, M. M. Cortical plasticity: from synapses to maps. *Annu. Rev. Neurosci.* **21**, 149–186 (1998).
- Sanes, J. N. & Donoghue, J. P. Plasticity and primary motor cortex. *Annu. Rev. Neurosci.* **23**, 393–415 (2000).
- Monfils, M. H., Plautz, E. J. & Kleim, J. A. In search of the motor engram: motor map plasticity as a mechanism for encoding motor experience. *Neuroscientist* **11**, 471–483 (2005).
- Nudo, R. J., Milliken, G. W., Jenkins, W. M. & Merzenich, M. M. Use-dependent alterations of movement representations in primary motor cortex of adult squirrel monkeys. *J. Neurosci.* **16**, 785–807 (1996).
- Kleim, J. A., Barbay, S. & Nudo, R. J. Functional reorganization of the rat motor cortex following motor skill learning. *J. Neurophysiol.* **80**, 3321–3325 (1998).
- Nudo, R. J., Wise, B. M., SiFuentes, F. & Milliken, G. W. Neural substrates for the effects of rehabilitative training on motor recovery after ischemic infarct. *Science* **272**, 1791–1794 (1996).
- Raineteau, O. & Schwab, M. E. Plasticity of motor systems after incomplete spinal cord injury. *Nature Rev. Neurosci.* **2**, 263–273 (2001).

8. Hebb, D. O. *The Organization of Behaviour: a Neuropsychological Theory* (Wiley, New York, 1949).
9. Iriki, A., Pavlides, C., Keller, A. & Asanuma, H. Long-term potentiation in the motor cortex. *Science* **245**, 1385–1387 (1989).
10. Hess, G., Aizenman, C. D. & Donoghue, J. P. Conditions for the induction of long-term potentiation in layer II/III horizontal connections of the rat motor cortex. *J. Neurophysiol.* **75**, 1765–1778 (1996).
11. Racine, R. J., Chapman, C. A., Trepel, C., Teskey, G. C. & Milgram, N. W. Post-activation potentiation in the neocortex. IV. Multiple sessions required for induction of long-term potentiation in the chronic preparation. *Brain Res.* **702**, 87–93 (1995).
12. Nudo, R. J., Jenkins, W. M. & Merzenich, M. M. Repetitive microstimulation alters the cortical representation of movements in adult rats. *Somatosens. Mot. Res.* **7**, 463–483 (1990).
13. Monfils, M. H., VandenBerg, P. M., Kleim, J. A. & Teskey, G. C. Long-term potentiation induces expanded movement representations and dendritic hypertrophy in layer V of rat sensorimotor neocortex. *Cereb. Cortex* **14**, 586–593 (2004).
14. Shors, T. J. & Matzel, L. D. Long-term potentiation: what's learning got to do with it? *Behav. Brain Sci.* **20**, 597–614 (1997).
15. Markram, H., Lubke, J., Frotscher, M. & Sakmann, B. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* **275**, 213–215 (1997).
16. Dan, Y. & Poo, M. Spike timing-dependent plasticity of neural circuits. *Neuron* **44**, 23–30 (2004).
17. Mavoori, J., Jackson, A., Diorio, C. & Fetz, E. E. An autonomous implantable computer for neural recording and stimulation in unrestrained primates. *J. Neurosci. Methods* **148**, 71–77 (2005).
18. Jackson, A., Moritz, C. T., Mavoori, J., Lucas, T. H. & Fetz, E. E. The Neurochip BCI: Towards a neural prosthesis for upper limb function. *IEEE Trans. Neural Sys. Rehab. Eng.* **14**, 187–190 (2006).
19. Jackson, A., Mavoori, J. & Fetz, E. E. Correlations between the same motor cortex cells and arm muscles during a trained task, free behaviour and natural sleep in the macaque monkey. *J. Neurophysiol.* (in the press).
20. Maynard, M. E. *et al.* Neuronal interactions improve cortical population coding of movement direction. *J. Neurosci.* **19**, 8083–8093 (1999).
21. Ito, M. & Kano, M. Long-lasting depression of parallel fibre-Purkinje cell transmission induced by conjunctive stimulation of parallel fibres and climbing fibres in the cerebellar cortex. *Neurosci. Lett.* **33**, 253–258 (1982).
22. Huang, Y.-Y., Pittenger, C. & Kandel, E. R. A form of long-lasting, learning-related synaptic plasticity in the hippocampus induced by heterosynaptic low-frequency pairing. *Proc. Natl Acad. Sci. USA* **101**, 859–864 (2004).
23. Baranyi, A. & Feher, O. Synaptic facilitation requires paired activation of convergent pathways in the neocortex. *Nature* **290**, 413–415 (1981).
24. Jankowska, E., Padel, Y. & Tanaka, R. The mode of activation of pyramidal tract cells by intracortical stimuli. *J. Physiol. (Lond.)* **249**, 617–636 (1975).
25. Jackson, A., Gee, V. J., Baker, S. N. & Lemon, R. N. Synchrony between neurons with similar muscle fields in monkey motor cortex. *Neuron* **38**, 115–125 (2002).
26. Fetz, E. E., Toyama, K. & Smith, W. A. in *Cerebral Cortex* Vol. 9 (ed. Peters, A.) 1–47 (Plenum, New York, 1991).
27. Hatsopoulos, N. G., Ojakangas, C. L., Paninski, L. & Donoghue, J. P. Information about movement direction obtained from synchronous activity of motor cortical neurons. *Proc. Natl Acad. Sci. USA* **95**, 15706–15711 (1998).
28. Fregnac, Y., Shulz, D., Thorpe, S. & Bienenstock, E. A cellular analogue of visual cortical plasticity. *Nature* **333**, 367–370 (1988).
29. Ahissar, E. *et al.* Dependence of cortical plasticity on correlated activity of single neurons and on behavioral context. *Science* **257**, 1412–1415 (1992).
30. Mu, Y. & Poo, M. Spike timing-dependent LTP/LTD mediates visual experience-dependent plasticity in a developing retinotectal system. *Neuron* **50**, 115–125 (2006).
31. Stickgold, R. & Walker, M. P. Sleep and memory: the ongoing debate. *Sleep* **28**, 1225–1226 (2005).
32. Prochazka, A., Mushahwar, V. K. & McCreery, D. B. Neural prostheses. *J. Physiol. (Lond.)* **533**, 99–109 (2001).
33. Plautz, E. J. *et al.* Post-infarct cortical plasticity and behavioral recovery using concurrent cortical stimulation and rehabilitative training: a feasibility study in primates. *Neurol. Res.* **25**, 801–810 (2003).
34. Khedr, E. M., Ahmed, M. A., Fathy, N. & Rothwell, J. C. Therapeutic trial of repetitive transcranial magnetic stimulation after acute ischemic stroke. *Neurology* **65**, 466–468 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Kirby for technical help, F. Miles for development of the Neurochip stimulator, L. Shupe for programming, and S. Perlmutter for advice. This work was supported by grants from the National Institutes of Health, the Office of Naval Research and the University of Washington Royalty Research Fund.

Author Contributions A.J. and E.E.F. conceived and designed the experiment. J.M. designed the Neurochip electronics. A.J. and J.M. performed the experiments. A.J. and E.E.F. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.E.F. (fetz@u.washington.edu).

Inactivation of the p53 pathway in retinoblastoma

Nikia A. Laurie^{1*}, Stacy L. Donovan^{1*}, Chie-Schin Shih¹, Jiakun Zhang¹, Nicholas Mills^{2,6}, Christine Fuller³, Amina Teunisse⁷, Suzanne Lam⁷, Yolande Ramos⁷, Adithi Mohan¹, Dianna Johnson⁸, Matthew Wilson^{3,4,8}, Carlos Rodriguez-Galindo⁵, Micaela Quarto⁹, Sarah Francoz¹⁰, Susan M. Mendrysa¹¹, R. Kiplin Guy², Jean-Christophe Marine¹⁰, Aart G. Jochemsen⁷ & Michael A. Dyer^{1,8}

Most human tumours have genetic mutations in their Rb and p53 pathways, but retinoblastoma is thought to be an exception. Studies suggest that retinoblastomas, which initiate with mutations in the gene retinoblastoma 1 (*RB1*), bypass the p53 pathway because they arise from intrinsically death-resistant cells during retinal development. In contrast to this prevailing theory, here we show that the tumour surveillance pathway mediated by Arf, MDM2, MDMX and p53 is activated after loss of *RB1* during retinogenesis. *RB1*-deficient retinoblasts undergo p53-mediated apoptosis and exit the cell cycle. Subsequently, amplification of the *MDMX* gene and increased expression of MDMX protein are strongly selected for during tumour progression as a mechanism to suppress the p53 response in *RB1*-deficient retinal cells. Our data provide evidence that the p53 pathway is inactivated in retinoblastoma and that this cancer does not originate from intrinsically death-resistant cells as previously thought. In addition, they support the idea that MDMX is a specific chemotherapeutic target for treating retinoblastoma.

Tumorigenesis involves sequential genetic lesions in pathways that regulate biological processes such as cell proliferation and cell survival^{1,2}. It has been proposed that both the p16^{Ink4a}-CycD/Cdk4-pRb and Arf-MDM2/MDMX-p53 pathways must be inactivated during tumorigenesis³. The primary role of the Rb pathway is to regulate cell proliferation^{3,4}, and that of the p53 pathway is to regulate responses to cellular insults (such as DNA damage or oncogenic stress)⁵⁻⁷. These pathways may be inactivated by mutations in their respective tumour suppressor genes, *RB1* and *p53* (also known as *TP53*) or in genes encoding modulators and/or effectors in these pathways. For example, some cancers have *MDM2* gene amplifications that suppress the p53 pathway by reducing steady-state amounts of the p53 protein⁸⁻¹⁰. When MDM2-mediated destabilization of p53 is blocked by the inhibitor nutlin-3 in tumours with *MDM2* gene amplifications, the p53 pathway is restored and tumour cells undergo p53-mediated cell-cycle arrest, cell death, or both^{11,12}. Therefore, identifying genetic perturbations in the Rb and p53 pathways can provide chemotherapeutic targets.

Retinoblastomas that arise from cells that have lost *RB1* have not been found to contain subsequent genetic lesions in the *p53* gene¹³ or p53 pathway¹⁴. Studies on *Rb;p107*-deficient mouse retinae have led to the proposal that retinoblastoma is a unique tumour that bypasses the p53 pathway because its cell of origin is intrinsically resistant to death^{15,16}. This theory has important implications for cancer genetics and treatment. It suggests that, depending on the cell of origin, cancer can proceed down a 'fast track' of tumorigenesis, because the cells are programmed to bypass some tumour suppressor pathways¹⁶. If this is true, then therapeutic targets may differ, depending on the initiating

genetic lesion and the pathways bypassed. Specifically, if the retinoblastoma cell of origin bypasses the p53 pathway, then chemotherapy targeting that pathway is inappropriate. Alternatively, if *p53* and downstream targets are intact and functional, then therapy that induces a p53 response may be effective.

***RB1* loss induces *p14*^{ARF} in human retinae**

A key component of the p53 tumour surveillance pathway is *p14*^{ARF} (ref. 3). When Rb activity is lost, the transcription factor E2F activates transcription of *p14*^{ARF} (ref. 17); *p14*^{ARF} then inactivates MDM2 (ref. 18), leading to p53-mediated apoptosis and exit from the cell cycle. If retinoblastomas arise from intrinsically death-resistant cells, then tumour cells with genetic perturbations that inactivate the p53 pathway will have no growth advantage. To determine whether the Arf-MDM2/MDMX-p53 oncogenic stress response pathway is intact in retinoblastoma, we isolated RNA and genomic DNA from human retinoblastomas. Expression of *p14*^{ARF} was increased 71- to 500-fold in the tumour samples as compared with normal human fetal retinae (Fig. 1a). Similar analyses of mouse retinoblastomas¹⁹ showed a 74- to 430-fold induction of *p19*^{Arf} expression (data not shown).

We acutely knocked down the expression of *RB1* in fetal week 14 (FW14) primary human retinae using an *RB1* siRNA²⁰ and found that *p14*^{ARF} was induced (Supplementary Fig. 1a-g). Similar results were obtained using a Cre-expressing plasmid in *Rb*^{Lox/Lox}; *p107*^{-/-} mouse retinae (data not shown). These data suggest that loss of *RB1* in the developing human retina or loss of *Rb* and *p107* in mouse retinae causes derepression of *Arf* and activation of the tumour surveillance mechanism.

¹Department of Developmental Neurobiology, ²Department of Chemical Biology and Therapeutics, ³Department of Pathology, ⁴Department of Surgery, Division of Ophthalmology, and ⁵Department of Hematology-Oncology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. ⁶Department of Cellular and Molecular Pharmacology, UCSF, San Francisco, California 94143, USA. ⁷Department of Molecular and Cell Biology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands. ⁸Department of Ophthalmology, University of Tennessee Health Science Center, Memphis, Tennessee 38163, USA. ⁹FIRC Institute of Molecular Oncology, 20139 Milan, Italy. ¹⁰Laboratory for Molecular Cancer Biology, Flanders Interuniversity Institute for Biotechnology, B-9052 Ghent, Belgium. ¹¹Basic Medical Sciences, Purdue University, West Lafayette, Indiana 47907, USA.

*These authors contributed equally to this work.

MDMX is amplified in retinoblastoma

Analysis by bacterial artificial chromosome comparative genome hybridization (BAC-CGH) showed that *MDMX*²¹, a gene related to *MDM2*, was amplified in three of seven fresh retinoblastoma samples; this amplification correlated with an increase in *MDMX* messenger RNA (Fig. 1b) and protein (Fig. 1c). We extended these data to include 49 paraffin-embedded retinoblastoma samples and carried out fluorescent *in situ* hybridization (FISH) analysis for *MDMX* and *MDM2* and immunohistochemistry for p53, p21, *MDM2* and *MDMX* (Fig. 1d–k and Supplementary Fig. 1h).

The *MDMX* gene copy number and the proportion of p53 ($t = -0.3321$; $P = 0.0096$) and p21 ($t = -0.2565$; $P = 0.0447$) immunopositive cells were inversely correlated, similar to previous studies in human breast tumours²². Of note, 32 of 49 (65%) human retinoblastomas had extra copies of *MDMX*, and 5 of 49 (10%) had extra copies of *MDM2* (Supplementary Table 1).

Functional p53 pathway downstream of MDMX

Perturbations of one gene in the p53 pathway relieves the selective pressure to inactivate other genes in the same pathway². For example, a cell line (Rh18) with an *MDM2* gene amplification has wild-type p53 and shows a robust p53 response to 5 Gy of ionizing radiation²³ (Supplementary Fig. 2). To test whether the p53 pathway is functional downstream of *MDMX* in retinoblastoma cells, we exposed Weri1 and Y79 human retinoblastoma cell lines (Supplementary Fig. 3) to 5 Gy of ionizing radiation^{24,25}. We used ML-1 leukaemia cells with wild-type p53 as a positive control²⁶, and the p53-deficient mouse retinoblastoma cell line SJMRBL-8 as a negative control (N.L. and M.A.D., unpublished data). After irradiation, the Y79, Weri1 and ML-1 cells showed an increase in p53 protein, phosphorylation of

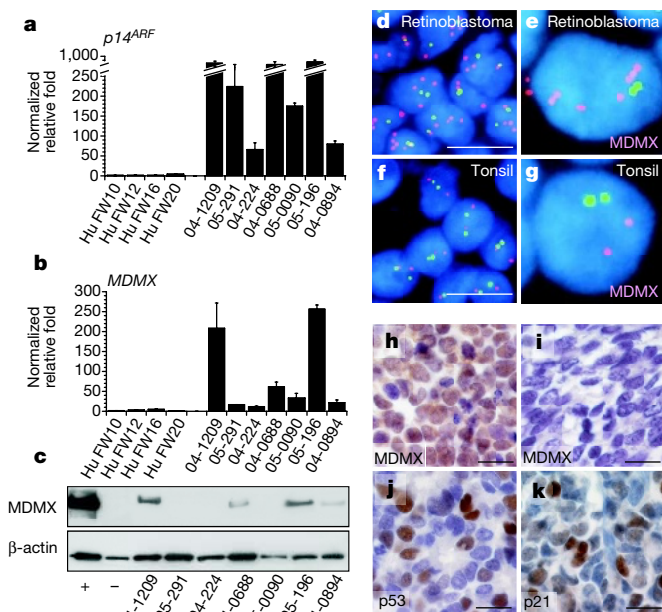


Figure 1 | MDMX amplification correlates with decreased activity of the p53 pathway. **a, b**, Real-time RT-PCR analysis of *p14^{ARF}* and *MDMX* expression was done on normal human fetal retinae at four stages of development and on seven retinoblastomas. Data from duplicate samples were normalized to *GAPDH* expression and are plotted as fold changes relative to that of FW10 human retinae (1.0-fold). Error bars represent the s.d. of two experiments. **c**, Samples analysed in **a** and **b** were immunoblotted for MDMX and β -actin. **d–g**, FISH analysis for *MDMX* was done on 49 primary untreated retinoblastomas; tonsil was used as a normal diploid control. Blue fluorescence is the nuclear counterstain; green fluorescence is the internal chromosomal control; red fluorescence is *MDMX*. **h–k**, The same 49 retinoblastomas were immunostained for MDMX, *MDM2*, p53 and p21. A retinoblastoma sample lacking *MDMX* was used as a negative control (**i**). Scale bars, 10 μ m.

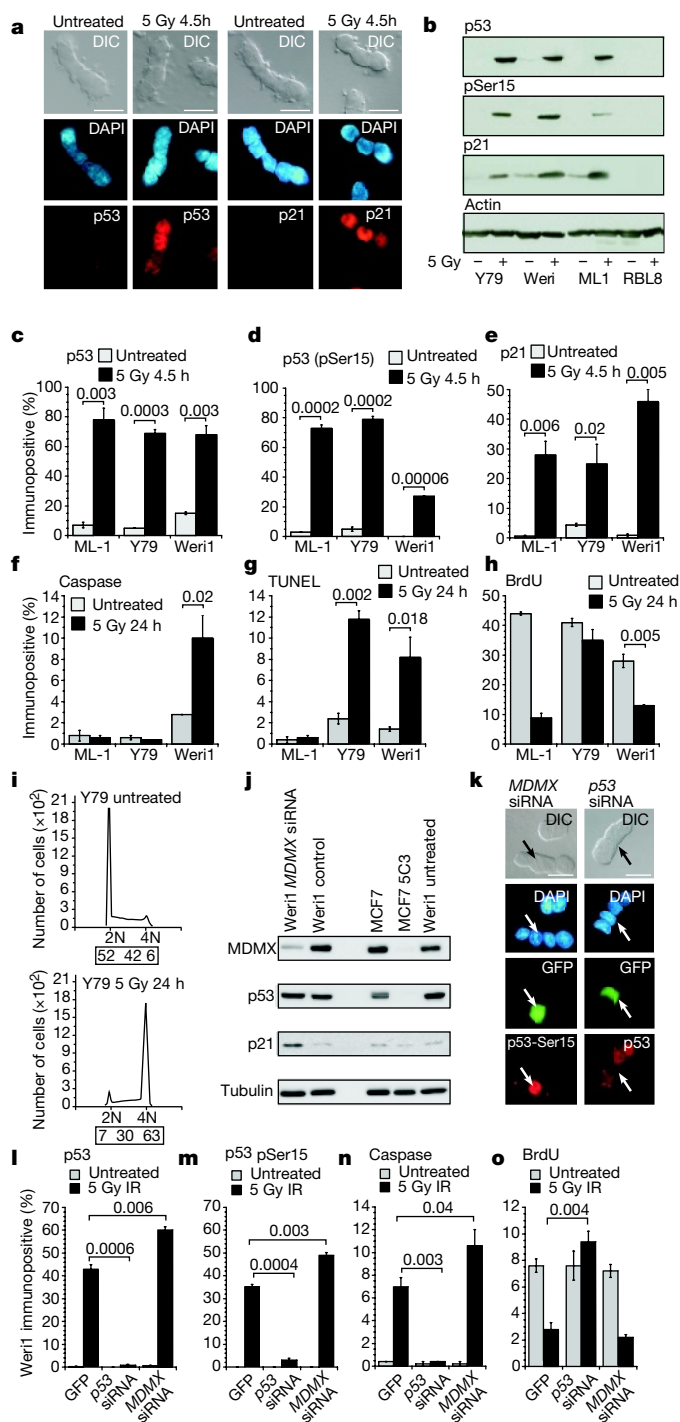


Figure 2 | Retinoblastoma cells show a p53 response to DNA damage. **a–h**, Weri1, Y79, ML-1 and SJMRBL-8 cells were exposed to 5 Gy of ionizing radiation. Immunofluorescence (**a**), immunoblotting (**b**) and dissociated cell scoring (**c–e**) confirmed that Weri1, Y79 and ML-1 cells induce the p53 pathway 4.5 h after ionizing radiation (IR). p53-deficient SJMRBL-8 cells did not activate the p53 pathway (**b**). **f–i**, ML-1, Y79 and Weri1 cells exited the cell cycle and/or initiated apoptosis 24 h after exposure to ionizing radiation. **j**, Weri1 cells transfected with the *MDMX* siRNA showed a reduction in *MDMX* protein and induction of the p53 target p21. **k–o**, Weri1 cells transfected with the p53 siRNA showed loss of p53 protein expression and very few p53⁺ cells in response to ionizing radiation. Cells transfected with the *MDMX* siRNA showed a more robust p53 response to ionizing radiation. In **c–h** and **l–o**, the proportion of immunopositive cells were scored in duplicate (250 cells each) from independent experiments. Error bars represent the s.d. of three (**c–h**) or two (**l–o**) experiments. *P* values are shown above the relevant bars. DIC, differential interference contrast microscopy. Scale bars, 10 μ m.

p53 on Ser 15, accumulation of p53 targets p21 and MDM2 (Fig. 2a–e and Supplementary Fig. 4a–e), cell-cycle exit and apoptosis (Fig. 2f–i and Supplementary Fig. 4f, g). Fresh retinoblastoma tumour cells from untreated enucleated eyes showed a similar p53 response to irradiation (Supplementary Fig. 5).

We next tested whether the response to ionizing radiation was p53 dependent and whether MDMX modulates the p53 pathway in retinoblastoma. Y79 and Weri1 cells were transfected with a vector encoding short interfering RNAs (siRNAs) targeted to p53 or MDMX^{22,27} (Fig. 2j) and 48 h later were exposed to 5 Gy of ionizing radiation. Cells transfected with the p53 siRNA contained fewer activated caspase-3-positive cells, fewer TdT-mediated dUTP nick end labelling (TUNEL)-positive cells, and fewer fragmented nuclei characteristic of late-stage apoptosis 24 h after exposure to 5 Gy of ionizing radiation (Fig. 2k–n and Supplementary Fig. 4h, i). Cells transfected with the p53 siRNA also contained more 5-bromodeoxyuridine (BrdU)-positive cells (Fig. 2o). Conversely, cells expressing the MDMX siRNA had a similar or more robust response to 5 Gy of ionizing radiation than did the controls (Fig. 2k–o and Supplementary Fig. 4h, i). Similar results were obtained with lentiviral vectors expressing MDMX and p53 siRNAs (Supplementary Fig. 6).

To confirm that the p53 pathway is intact downstream of MDMX in retinoblastoma cells, we ectopically expressed p53, which has been shown to elicit a robust p53 response in p53-null cells but not in wild-type cells^{23,28}. Neither cell proliferation nor viability was altered after ectopic p53 expression in Weri1 or Y79 cells (Supplementary Fig. 4j–l).

MDMX promotes retinoblastoma in mice

Inactivation of *Rb* and *p107* can lead to retinoblastoma in chimaeric mice²⁹, and *p107*-deficient mice with *Rb* deletion targeted to the

developing retina are susceptible to retinoblastoma^{15,19,30}; however, the penetrance is low, tumour progression is slow, and the tumours are not as aggressive or invasive as human retinoblastomas^{31,32}. By contrast, mice lacking *p107*, *Rb* and *p53* develop 100% penetrant bilateral retinoblastoma that is aggressive and invasive^{19,31,32}. Although these data indicate that p53 has a role in retinoblastoma, they do not recapitulate the precise genetic changes that occur in human retinoblastomas, which express wild-type p53.

If increased MDMX expression contributes to tumorigenesis, then ectopic expression of MDMX in *Rb*; *p107*-deficient retinoblastoma should promote tumour progression similar to that observed in *Chx10*–*Cre*; *Rb*^{Lox/Lox}; *p107*^{−/−}; *p53*^{Lox/Lox} mice³². To test this hypothesis, we used square-wave electroporation to introduce a plasmid expressing Cre recombinase or Cre recombinase and MDMX (Supplementary Fig. 7a) into the eyes of newborn *Rb*^{Lox/Lox}; *p107*^{−/−} pups. The survival and proliferation of transfected cells were analysed 7 and 14 d after electroporation. Expression of MDMX promoted proliferation and survival in developing retinal cells lacking *Rb* and *p107* (Fig. 3a–c). Moreover, these cells expressed the retinal progenitor cell marker Pax6, which is expressed in mouse retinoblastomas^{19,20} (Fig. 3a, d).

We then injected a plasmid expressing MDMX and an alkaline phosphatase reporter gene into the subretinal space of newborn *Pax6*–*Cre*; *Rb*^{Lox/Lox}; *p107*^{−/−} pups and subsequently electroporated it into the developing retinal cells. After 3 weeks, the retinae were isolated and stained for alkaline phosphatase expression. *Rb* is reported to be inactivated in the peripheral 30–40% of the *Pax6*–*Cre*; *Rb*^{Lox/Lox}; *p107*^{−/−} retinae³³; therefore, within a single retina, we compared the effects of ectopic MDMX expression in cells lacking *Rb* and *p107* (peripheral retina) with those in cells lacking only *p107* (central retina).

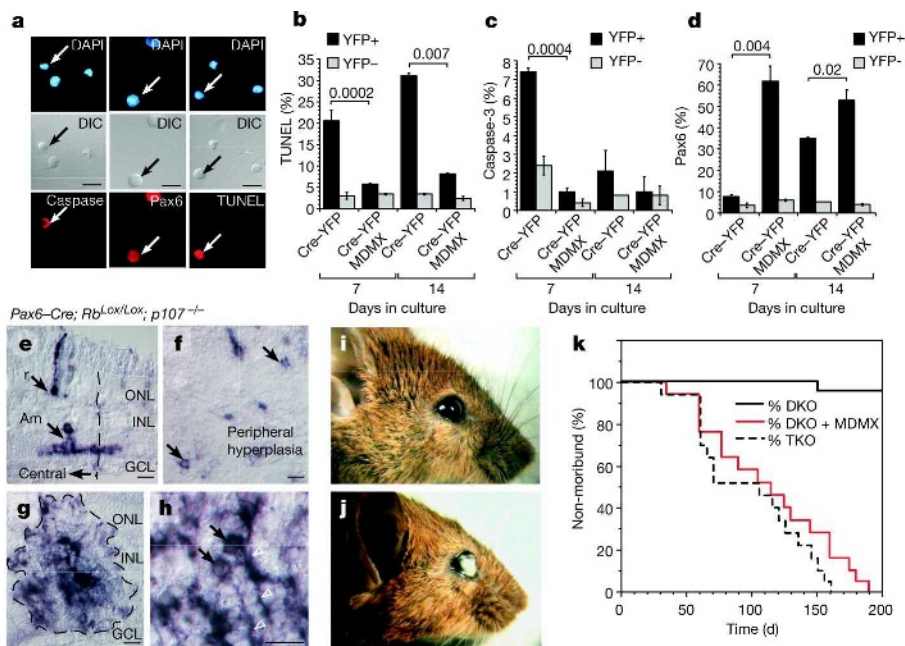


Figure 3 | MDMX promotes retinal tumorigenesis in mice. a–d, Postnatal day 0 (P0) *Rb*^{Lox/Lox}; *p107*^{−/−} retinae were square-wave electroporated with a plasmid expressing Cre and a yellow fluorescent protein (YFP) reporter gene or Cre, a YFP reporter gene and a Flag-tagged MDMX cDNA. Dissociated cell scoring of FACS-purified YFP⁺ and YFP[−] cells was done 7 and 14 d later. b–d, Cells (*n* = 250) were scored for each sample in duplicate. Error bars represent the s.d. of two experiments. e–h, The MDMX cDNA and an alkaline phosphatase (AP) reporter gene were square-wave electroporated into P0 *Pax6*–*Cre*; *Rb*^{Lox/Lox}; *p107*^{−/−} retinae. e, Representative examples of normal rod photoreceptors (r) and amacrine cells (Am) in the central retina where Pax6–Cre is not active. f, Single alkaline-phosphatase-labelled cells were evident in hyperplastic peripheral retinae when the control alkaline

phosphatase plasmid was electroporated into *Pax6*–*Cre*; *Rb*^{Lox/Lox}; *p107*^{−/−} eyes. g, h, By contrast, ectopic MDMX expression promoted expansion of cells (arrows) that extended processes (arrowheads) characteristic of retinoblastomas in *Chx10*–*Cre*; *Rb*^{Lox/Lox}; *p107*^{−/−}; *p53*^{Lox/Lox} mice. i–k, The tumours that arose from ectopic expression of MDMX were aggressive and invasive and led to an increase in the proportion of moribund mice (j) as compared with *Chx10*–*Cre*; *Rb*^{Lox/Lox}; *p107*^{−/−} mice. DIC, differential interference contrast microscopy; ONL, outer nuclear layer; INL, inner nuclear layer; GCL, ganglion cell layer. DKO, *Chx10*–*Cre*; *Rb*^{Lox/Lox}; *p107*^{−/−}; TKO, *Chx10*–*Cre*; *Rb*^{Lox/Lox}; *p107*^{−/−}; *p53*^{Lox/Lox}. P values are shown above the relevant bars. Scale bars, 10 μ m.

Ectopic expression of *MDMX* in the central retina had little effect on proliferation or differentiation (Fig. 3e). Some hyperplasia formed in the periphery of *Pax6-Cre;Rb^{Lox/Lox};p107^{-/-}* retinæ, but we observed minimal clonal expansion of individual cells in that region

(Fig. 3f). When *MDMX* was expressed in cells lacking *Rb* and *p107* in the peripheral retina that lacked hyperplasia, clonal expansion occurred (Fig. 3g, h). Moreover, those cells showed morphological features (Supplementary Fig. 7b–j and Supplementary Tables 2–4) and aggressive invasive retinoblastoma similar to those of *Chx10-Cre;Rb^{Lox/Lox};p53^{Lox/Lox};p107^{-/-}* mice (Fig. 3i–k).

MDMX promotes human retinoblastoma

We electroporated primary human FW14 retinæ with an *RB1* siRNA, an *MDMX* cDNA and a green fluorescent protein (GFP) reporter gene. A mutant form of *MDMX* (*MDMX-G57A*) that cannot bind to p53 (ref. 22) was used as a control (Supplementary Fig. 7k). Cells electroporated with a control siRNA differentiated and extended processes after 10 d in explant culture (Fig. 4a). Those electroporated with the *RB1* siRNA underwent extensive apoptosis (Fig. 4b, c). When the *RB1* siRNA and an *MDMX* cDNA were electroporated together, by contrast, minimal cell death occurred and the immature cells organized into rosettes similar to retinoblastoma (Fig. 4d, e). The *MDMX-G57A* mutant confirmed that suppression of cell death was specific to the p53 pathway (Fig. 4f).

To quantify changes in proliferation and cell survival, we repeated the above experiment with modifications. After 5 d in culture, the retinæ were treated with [³H]thymidine for 24 h to label all proliferating cells. After another 48 h, they were treated with a 1-h pulse of BrdU, again to label proliferating cells. The proportion of BrdU-positive cells was significantly greater when *MDMX* was coexpressed with the *RB1* siRNA but not when *MDMX-G57A* was expressed; similar results were observed in the [³H]thymidine-positive cells and the double-positive cells (Fig. 4g, h), which continued to divide 5–7 d in culture.

A higher proportion of Pax6-expressing cells were found among the *RB1*-deficient retinoblasts (Supplementary Fig. 7l) expressing *MDMX*, consistent with their immature cell morphology (Fig. 4d). In addition, *MDMX* promoted the progression from differentiated early stage tumour cells that resemble amacrine and/or horizontal cells to tumour cells with features of less differentiated cells (Supplementary Fig. 7m–q and Supplementary Tables 5–7). By scoring the proportions of activated caspase-3-positive cells and TUNEL-positive cells, we confirmed that *MDMX* blocked cell death in *Rb*-deficient human retinoblasts through its ability to bind and inactivate p53, as indicated by the inability of *MDMX-G57A* to replicate this action (Fig. 4i).

Nutlin-3 blocks MDMX in retinoblastoma

Computational modelling of nutlin-3, a small-molecule inhibitor of the MDM2–p53 interaction¹² (Fig. 5a, b), suggested that nutlin-3 may block the MDMX–p53 interaction (Fig. 5c, d). *MDMX* and *MDM2* bind p53 (ref. 34) with similar affinities (dissociation constant, $K_d = 0.5 \mu\text{M}$; Fig. 5e). Racemic nutlin-3 bound to *MDM2* with an inhibition constant (K_i) of $0.7 \mu\text{M}$ (Fig. 5f), confirming published results using enantiomerically pure nutlin-3a¹². Racemic nutlin-3 also specifically bound to *MDMX* with a K_i of $28 \mu\text{M}$ ($\sim 14 \mu\text{M}$ inferred K_i for nutlin-3a; Fig. 5f).

Coimmunoprecipitation experiments showed that nutlin-3 prevents the MDMX–p53 interaction in cells (Fig. 5g and Supplementary Fig. 8a), and a mouse embryonic fibroblast (MEF) growth assay^{22,35} showed that nutlin-3 can induce the p53 pathway by inhibiting MDMX binding to p53 (Supplementary Fig. 8c, d). Of note, retinoblastoma cells (Weri1) with wild-type p53 and *MDMX* amplification (Supplementary Fig. 3) were sensitive to racemic nutlin-3 (50% limiting concentration $\text{LC}_{50} = 0.7 \mu\text{M}$), whereas a p53-deficient retinoblastoma cell line (SJMRBL-8) was insensitive (Fig. 5h). By altering the quantities of p53 or *MDMX* in Weri1 retinoblastoma cells, we could change the sensitivity to nutlin-3 (Supplementary Fig. 8e–g).

To test whether the mechanism by which nutlin-3 induces a p53 response in retinoblastoma cells with *MDMX* amplification is

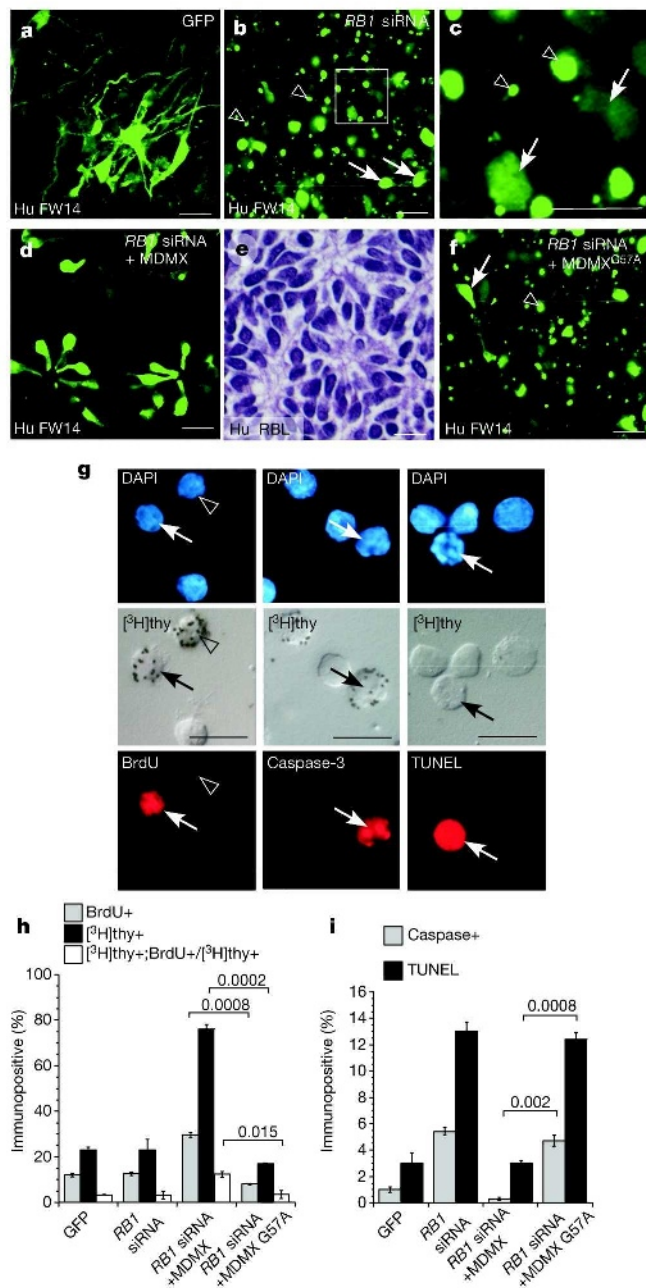


Figure 4 | MDMX rescues cell death in *RB1*-deficient human retinoblasts. FW14 human retinæ were square-wave electroporated with a GFP reporter gene along with the indicated expression vectors and cultured for 10 d as explants. **a**, Normal differentiated GFP⁺ neurons were readily identified in the control sample. **b, c**, A few immature cells (arrows) persisted after introduction of the *RB1* siRNA, but most of the cells died (arrowheads). **d, e**, Co-electroporation of the *MDMX* cDNA with the *RB1* siRNA blocked cell death and led to rosette formation similar to human retinoblastoma. **f**, The *MDMX-G57A* mutant confirmed that this effect was specific to the p53 pathway. **g–i**, After 5 d in culture, human fetal retinæ electroporated with the indicated plasmids were labelled with [³H]thymidine for 24 h, and 48 h later were labelled with BrdU for 1 h. **h**, The proportions of BrdU⁺, [³H]thymidine⁺ and double-positive cells were scored in duplicate. **i**, The proportions of activated caspase-3⁺ and TUNEL⁺ cells were also scored. Error bars in **h** and **i** represent the s.d. from two independent samples. *P* values are shown above the relevant bars. Scale bars, 10 μm .

through MDM2 rather than through direct binding, we carried out experiments in *Mdm2*-deficient MEFs (Supplementary Figs 9 and 10 and Table 8). MdmX partially rescued the loss of Mdm2 in MEFs, and the effect of MdmX was blocked by nutlin-3 even in the absence of Mdm2 (Supplementary Figs 9 and 10).

Several preclinical models of retinoblastoma have been developed and used to test combinations of broad-spectrum chemotherapy^{19,32,36}. One drug, topotecan, induces a p53 response in retinoblastoma cells that is similar to that induced by 5 Gy of ionizing radiation (Supplementary Fig. 8h, i). Nutlin-3 also induces a p53 response in these cells (Supplementary Fig. 8j–n). The combination of topotecan and nutlin-3 synergistically killed retinoblastoma cells in culture (Supplementary Fig. 8o) and in our preclinical retinoblastoma models³⁶ (Fig. 5i–m). Combined subconjunctival injection of topotecan and nutlin-3 reduced tumour burden 82-fold with no systemic or ocular side-effects. The best combination of systemic chemotherapy in this model resulted in only a fivefold reduction in tumour burden, and the mice suffered from severe complications associated with systemic broad-spectrum chemotherapy³⁶. Even though nutlin-3 binds less efficiently to MDMX than to MDM2, intraocular concentrations of nutlin-3 achieved by subconjunctival injection should be sufficient to disrupt both MDM2–p53 and MDMX–p53 in retinoblastomas.

Discussion

A previous study correlated apoptosis with retinal-cell-type markers and concluded that, in mice, retinoblastoma originates from intrinsically death-resistant cells^{15,16}. Because human retinoblastomas express wild-type p53, it was assumed that the p53 pathway was intact and the status of the other genes in the pathway (such as *p14^{ARF}*, *MDM2* and *MDMX*) was not considered.

Inactivation of the Rb pathway in the developing mouse or human retina leads to ectopic proliferation and activation of the Arf–MDM2/MDMX–p53 tumour surveillance pathway. The ectopic proliferation caused by the loss of the Rb pathway is balanced, to some extent, by p53-mediated apoptosis. Additional genetic changes (such as *MDMX* gene amplification) occur in the preneoplastic retinoblastoma cells, and cells in which the p53 pathway is inactivated have a growth advantage over those with an intact Arf–MDM2/MDMX–p53 tumour surveillance network. On the basis of these data, we propose that cells with disruptions in the Rb and p53 pathways clonally expand and form retinoblastoma. We propose that inactivation of the p53 pathway promotes the transition from differentiated retinoblastoma cells with amacrine/horizontal cell features to a more immature cell with retinal progenitor cell features. It is also possible, however, that there are two distinct tumour cell types (amacrine/horizontal and progenitor) that originate from distinct cells of origin¹⁶.

These findings not only challenge the long-standing belief that retinoblastoma is the exception to the rule that the Rb and p53 pathways must be inactivated in cancer, but also provide a specific target for chemotherapy. Nutlin-3 antagonizes the MDMX–p53 interaction and efficiently kills retinoblastoma cells. In addition, combining nutlin-3 with topotecan synergistically increases tumour cell killing. On the basis of our preclinical studies, we propose that subconjunctival administration of these two drugs could achieve the same synergistic effect in individuals with retinoblastoma without causing the side-effects associated with prolonged systemic exposure to broad-spectrum chemotherapeutic drugs. Local delivery of nutlin-3 or other MDMX inhibitors being developed may also be beneficial for the treatment of breast, colon, lung and prostate cancers with *MDMX* amplifications.

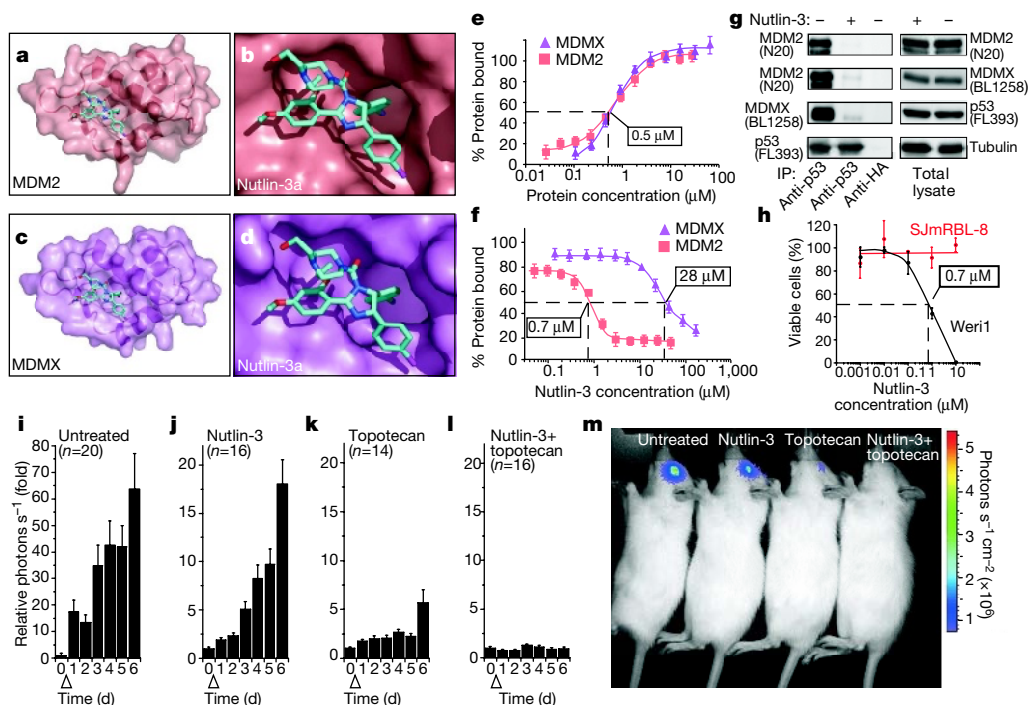


Figure 5 | Nutlin-3 inhibits MDMX activity in retinoblastoma. **a–d**, Three-dimensional modelling suggests that nutlin-3 can bind the p53 pocket of MDMX. **e**, Recombinant MDM2 and MDMX bind the p53 peptide with equal affinity (K_d = 0.5 μ M). **f**, Racemic nutlin-3 specifically competes for the p53 peptide bound to MDM2 and MDMX. The K_i for MDM2–p53 was 0.7 μ M, and that for MDMX–p53 was 28 μ M. Error bars in **e** and **f** represent the s.d. of four samples. **g**, Coimmunoprecipitation experiments show that 10 μ M racemic nutlin-3 reduced p53 binding of both MDM2 and MDMX.

h, Sensitivity of Weri1 (black line; LC_{50} = 0.7 μ M) and p53-deficient SJMRBL-8 (red line) cells to nutlin-3. All data were scored in duplicate; error bars represent the s.d. of three independent samples. **i–m**, Subconjunctival nutlin-3 and topotecan administered to an orthotopic retinoblastoma xenograft model for 5 d. Data are plotted relative to the initial luciferase activity for each mouse before administration of drug and represent the mean $\pm$ s.d. of the number of mice (14–20) indicated for each group.

METHODS

A detailed description of materials and methods is given in Supplementary Information. Individual protocols are provided on the authors' website (<http://www.stjude.org/dyer>).

Mouse and rat strains. *Rb*^{+/−} mice were obtained from The Jackson Laboratory; *p53*^{Lox/Lox} and *Rb*^{Lox/Lox} mice from the National Cancer Institute; *p107* knockout mice from T. Jacks; *Chx10-Cre* mice from C. Cepko; and *Pax6-Cre* mice from R. Ashery-Padan. All mice were crossed to C57Bl/6 mice purchased from Charles River Laboratories. Timed-pregnant Sprague Dawley rats were obtained from Charles River Laboratories. The xenograft model of retinoblastoma has been described³⁶.

Human retinae. Human retinae were obtained from Advanced Bioscience Resources. They were maintained in culture by using protocols that we previously developed for mouse retinal cultures.

Received 30 June; accepted 24 August 2006.

- Hahn, W. C. & Weinberg, R. A. Modelling the molecular circuitry of cancer. *Nature Rev. Cancer* **2**, 331–341 (2002).
- Vogelstein, B. & Kinzler, K. W. Cancer genes and the pathways they control. *Nature Med.* **10**, 789–799 (2004).
- Sherr, C. J. & McCormick, F. The RB and p53 pathways in cancer. *Cancer Cell* **2**, 103–112 (2002).
- Chau, B. N. & Wang, J. Y. Coordinated regulation of life and death by RB. *Nature Rev. Cancer* **3**, 130–138 (2003).
- Vogelstein, B., Lane, D. & Levine, A. J. Surfing the p53 network. *Nature* **408**, 307–310 (2000).
- Oren, M. Decision making by p53: life, death and cancer. *Cell Death Differ.* **10**, 431–442 (2003).
- Prives, C. & Hall, P. A. The p53 pathway. *J. Pathol.* **187**, 112–126 (1999).
- Honda, R., Tanaka, H. & Yasuda, H. Oncoprotein MDM2 is a ubiquitin ligase E3 for tumor suppressor p53. *FEBS Lett.* **420**, 25–27 (1997).
- Kubbutat, M. H., Jones, S. N. & Vousden, K. H. Regulation of p53 stability by Mdm2. *Nature* **387**, 299–303 (1997).
- Momand, J., Jung, D., Wilczynski, S. & Niland, J. The MDM2 gene amplification database. *Nucleic Acids Res.* **26**, 3453–3459 (1998).
- Yang, Y. *et al.* Small molecule inhibitors of HDM2 ubiquitin ligase activity stabilize and activate p53 in cells. *Cancer Cell* **7**, 547–559 (2005).
- Vassilev, L. T. *et al.* In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **303**, 844–848 (2004).
- Kato, M. V. *et al.* Loss of heterozygosity on chromosome 17 and mutation of the p53 gene in retinoblastoma. *Cancer Lett.* **106**, 75–82 (1996).
- Nork, T. M., Poulsen, G. L., Millecchia, L. L., Jantz, R. G. & Nickells, R. W. p53 regulates apoptosis in human retinoblastoma. *Arch. Ophthalmol.* **115**, 213–219 (1997).
- Chen, D. *et al.* Cell-specific effects of RB or RB/p107 loss on retinal development implicate an intrinsically death-resistant cell-of-origin in retinoblastoma. *Cancer Cell* **5**, 539–551 (2004).
- Dyer, M. A. & Bremner, R. The search for the retinoblastoma cell of origin. *Nature Rev. Cancer* **5**, 91–101 (2005).
- Aslanian, A., Iaquinia, P. J., Verona, R. & Lees, J. A. Repression of the Arf tumor suppressor by E2F3 is required for normal cell cycle kinetics. *Genes Dev.* **18**, 1413–1422 (2004).
- Lowe, S. W. & Sherr, C. J. Tumor suppression by Ink4a-Arf: progress and puzzles. *Curr. Opin. Genet. Dev.* **13**, 77–83 (2003).
- Zhang, J., Schweers, B. & Dyer, M. A. The first knockout mouse model of retinoblastoma. *Cell Cycle* **3**, 952–959 (2004).
- Donovan, S., Schweers, B., Martins, R., Johnson, D. & Dyer, M. A. Compensation by tumor suppressor genes during retinal development in mice and humans. *BMC Biol.* **4**, 14 (2006).
- Shvarts, A. *et al.* MDMX: a novel p53-binding protein with some functional properties of MDM2. *EMBO J.* **15**, 5349–5357 (1996).
- Danovi, D. *et al.* Amplification of Mdmx (or Mdm4) directly contributes to tumor formation by inhibiting p53 tumor suppressor activity. *Mol. Cell. Biol.* **24**, 5835–5843 (2004).
- McKenzie, P. P., McPake, C. R., Ashford, A. A., Vanin, E. F. & Harris, L. C. MDM2 does not influence p53-mediated sensitivity to DNA-damaging drugs. *Mol. Cancer Ther.* **1**, 1097–1104 (2002).
- Canman, C. E. *et al.* Activation of the ATM kinase by ionizing radiation and phosphorylation of p53. *Science* **281**, 1677–1679 (1998).
- Kastan, M. B., Lim, D. S., Kim, S. T. & Yang, D. ATM—a key determinant of multiple cellular responses to irradiation. *Acta Oncol.* **40**, 686–688 (2001).
- Kastan, M. B., Onyekwere, O., Sidransky, D., Vogelstein, B. & Craig, R. W. Participation of p53 protein in the cellular response to DNA damage. *Cancer Res.* **51**, 6304–6311 (1991).
- Brummelkamp, T. R., Bernards, R. & Agami, R. A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296**, 550–553 (2002).
- Baker, S. J., Markowitz, S., Fearon, E. R., Willson, J. K. & Vogelstein, B. Suppression of human colorectal carcinoma cell growth by wild-type p53. *Science* **249**, 912–915 (1990).
- Robanus-Maandag, E. *et al.* p107 is a suppressor of retinoblastoma development in pRb-deficient mice. *Genes Dev.* **12**, 1599–1609 (1998).
- MacPherson, D. *et al.* Cell type-specific effects of Rb deletion in the murine retina. *Genes Dev.* **18**, 1681–1694 (2004).
- Dyer, M. A. & Harbour, J. W. in *Clinical Ocular Oncology* Ch. 66 (eds Singh, A. D., Damato, B., Murphree, A. L. & Perry, J. D.) (Elsevier, London, in the press).
- Dyer, M. A., Rodriguez-Galindo, C. & Wilson, M. W. Use of preclinical models to improve treatment of retinoblastoma. *PLoS Med.* **2**, e332 (2005).
- Marquardt, T. *et al.* Pax6 is required for the multipotent state of retinal progenitor cells. *Cell* **105**, 43–55 (2001).
- Lu, F. *et al.* Proteomimetic libraries: design, synthesis, and evaluation of p53–MDM2 interaction inhibitors. *J. Comb. Chem.* **8**, 315–325 (2006).
- Dang, J. *et al.* The RING domain of Mdm2 can inhibit cell proliferation. *Cancer Res.* **62**, 1222–1230 (2002).
- Laurie, N. A. *et al.* Topotecan combination chemotherapy in two new rodent models of retinoblastoma. *Clin. Cancer Res.* **11**, 7569–7578 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Harris, G. Zambetti and M. Baron for discussions; S. Pounds for statistical analysis; M. Roussel for MDMX and MDM2 retroviruses; B. Schulman and D. Bashford for assistance with MDMX–nutlin-3 modelling; A. McArthur for editorial assistance; J. Gray for assistance with real-time RT–PCR and genomic DNA preparations; and F. Carlotti and M. Rabeling for advice on lentiviral experiments and production of lentivirus stocks. This work was supported by grants (to M.A.D.) from the National Eye Institute, Cancer Center Support from the National Cancer Institute, the American Cancer Society, Research to Prevent Blindness, the Pearle Vision Foundation, the International Retinal Research Foundation and the American Lebanese Syrian Associated Charities (ALSAC). M.A.D. is a Pew Scholar. This work was also supported by funding from the Association for International Cancer Research (A.G.J.) and EC FP6 (A.G.J. and J.-C.M.), the Dutch Cancer Society (Y.R.), the Belgian Foundation against Cancer (J.-C.M.) and Télévie (S.F.). This publication was supported in part by a grant from the National Cancer Institute.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.A.D. (michael.dyer@stjude.org).

Self-cooling of a micromirror by radiation pressure

S. Gigan^{1,2}, H. R. Böhm^{1,2}, M. Paternostro^{2,†}, F. Blaser², G. Langer³, J. B. Hertzberg^{4,5}, K. C. Schwab^{4,†}, D. Bäuerle^{3,†}, M. Aspelmeyer^{1,2} & A. Zeilinger^{1,2}

Cooling of mechanical resonators is currently a popular topic in many fields of physics including ultra-high precision measurements¹, detection of gravitational waves^{2,3} and the study of the transition between classical and quantum behaviour of a mechanical system^{4–6}. Here we report the observation of self-cooling of a micromirror by radiation pressure inside a high-finesse optical cavity. In essence, changes in intensity in a detuned cavity, as caused by the thermal vibration of the mirror, provide the mechanism for entropy flow from the mirror's oscillatory motion to the low-entropy cavity field². The crucial coupling between radiation and mechanical motion was made possible by producing free-standing micromirrors of low mass ($m \approx 400$ ng), high reflectance (more than 99.6%) and high mechanical quality ($Q \approx 10,000$). We observe cooling of the mechanical oscillator by a factor of more than 30; that is, from room temperature to below 10 K. In addition to purely photothermal effects⁷ we identify radiation pressure as a relevant mechanism responsible for the cooling. In contrast with earlier experiments, our technique does not need any active feedback^{8–10}. We expect that improvements of our method will permit cooling ratios beyond 1,000 and will thus possibly enable cooling all the way down to the quantum mechanical ground state of the micromirror.

Radiation-pressure forces inside optical cavities are known to pose an ultimate limit on the sensitivity of interferometric measurements^{11,12}. However, less well known is the fact that radiation pressure can also be used for the opposite, namely to counteract the dynamics of a cavity mirror by means of dynamical back action^{2,7,13}. In a recent experiment a passive cooling mechanism for a micromechanical oscillator based on bolometric back action was presented⁷. Even though this scheme has intrinsically limited cooling capability because it ultimately relies on heating by absorption, it might permit a quantitatively significant reduction of the oscillator's thermal motion. A more powerful scheme is provided by the use of radiation pressure as a feedback force². In this case, optical absorption does not impose a fundamental limit. The difficulty in using radiation pressure for this cooling purpose is that it calls for stable control of the detuning of a high-finesse cavity, strong optomechanical coupling and a low mass of at least one cavity mirror, thus requiring nanomechanical or micromechanical systems of high optical and mechanical quality (characterized by the cavity finesse F and the mechanical quality factor Q). Although cavity-induced radiation-pressure effects have already been used to modify elastic properties of mirrors^{7,14–16} and to enforce mechanical instabilities^{14,17–19}, none of the previous experiments was able to combine these strict requirements. We have overcome this limitation by developing a method to produce free-standing micromirrors of low mass (about 400 ng), high reflectance (more than 99.6%) and high mechanical quality ($Q \approx 10^4$). The use

of such micromirrors in a detuned optical cavity allows us to observe for the first time self-cooling in a regime in which, although photothermal effects are still present, radiation pressure participates significantly in the self-cooling process.

Radiation-pressure forces in an optical cavity arise from the momentum transfer of photons reflected from the mirror surface. For certain cavity detuning—that is, if the cavity angular frequency ω_c is off resonance with the frequency ω_l of the pump laser, the radiation pressure is highly sensitive to small displacements of the cavity mirror. This is a consequence of the fact that the energy stored in a cavity field varies strongly with detuning. As a consequence, the dynamics of an oscillating mirror inside a detuned cavity is modified by a mechanical rigidity that depends on the detuning. For a high-finesse cavity, the radiation-pressure-induced back action can act on the mirror motion in such a way as to induce low-noise damping. This is the general concept of dynamical back action¹³. A simple classical description of the dynamics of the mirror shows that both the resonance frequency ω_M and the natural damping rate γ of the mirror motion are modified by radiation pressure to ω_{eff} and γ_{eff} , respectively^{2,7}. In particular, within the classical framework, the modified damping rate follows

$$\gamma_{\text{eff}} = \gamma + \frac{\beta(\Delta)}{2m} \frac{2\kappa}{(2\kappa)^2 + \omega_M^2} \quad (1)$$

with the cavity decay rate $\kappa = \pi c/2FL$, the cavity finesse F , the cavity length L and the speed of light *in vacuo*, c . Optimum damping is achieved when $1/(2\kappa)$ is of the order of ω_M , which for ω_M in the megahertz range requires a high-finesse cavity. Equation (1) depends on $\beta(\Delta)$, the spatial gradient of the radiation force evaluated at a (spatial) detuning $\Delta_x = L\Delta/\omega_l$. Here Δ is the effective detuning between cavity and laser frequency, including the effect of radiation pressure²⁰. The contribution β , induced by radiation pressure, can be positive or negative depending on the sign of Δ . It is straightforward to show that $\beta(\Delta)$ is negative for $\Delta < 0$, corresponding to $\gamma_{\text{eff}} < \gamma$. In this regime, the system can enter instability. The focus of this work is the investigation of the opposite regime ($\beta(\Delta) > 0$) in which $\gamma_{\text{eff}} > \gamma$. This low-noise damping results in a reduction of the mirror temperature, and hence self-cooling is achieved. The previous self-cooling experiments based on bolometric forces⁷ were operated in the regime of negative detuning, where radiation pressure counteracts the cooling.

To observe the self-cooling effect a read-out scheme of the mirror motion is required. For this purpose it is sufficient to measure the statistical properties of the optical field that leaks out of the cavity. In a way, the output cavity field represents a 'blank sheet' on which the dynamics of the mirror can be written. It is possible to briefly sketch the main idea of our self-cooling read-out process by exploiting a

¹Physics Faculty, Institute for Experimental Physics, University of Vienna, Boltzmanngasse 5, A-1090 Vienna, Austria. ²Institute for Quantum Optics and Quantum Information (IQOQI), Austrian Academy of Sciences, Boltzmanngasse 3, A-1090 Vienna, Austria. ³Institute for Applied Physics, Johannes-Kepler-University Linz, Altenbergerstrasse 69, A-4040 Linz, Austria. ⁴Laboratory for Physical Sciences, University of Maryland, College Park, Maryland 20740, USA. ⁵Department of Physics, University of Maryland, College Park, Maryland 20740, USA. [†]Present addresses: School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, UK (M.P.); Cornell University, Ithaca, New York 14853, USA (K.C.S.); University of California, Berkeley, California 94720-1740, USA (D.B.).

simple (but for our purposes sufficient) semiclassical picture. A full quantum mechanical framework, which generalizes the classical picture for self-cooling proposed so far in the literature^{2,10}, is presented elsewhere²¹. Not only is this (more general) approach in agreement with the classical picture taken into account by equation (1) but it also paves the way towards the rigorous study of the limitations imposed on self-cooling by the influences of quantum noise²¹. The total energy of a cavity consisting of a fixed mirror and a movable mirror driven by an input laser field of power P is given by^{20,21}

$$U = \frac{\hbar}{2}(\omega_c - \omega_l)(X^2 + Y^2) - \hbar \frac{\omega_c}{2L}(X^2 + Y^2 - 1)q + \frac{1}{2}\left(\frac{p^2}{m} + m\omega_M^2 q^2\right) + \sqrt{2\hbar E}Y \quad (2)$$

where X and Y are the quadratures of the cavity field, p and q are the momentum and position quadratures of the oscillating mirror, and $E = \sqrt{2\kappa P/\hbar\omega_l}$ is the coupling rate between the cavity and the input laser field. If the timescale set by the cavity decay rate is the shortest in the dynamics of the system—that is, $\kappa \gg \omega_M$ —the cavity field follows the mirror motion adiabatically. As a consequence, the fluctuations δY_{out} of the field leaking out of the cavity are directly related to the fluctuations of the mirror's position quadrature as $\delta Y_{\text{out}}(t) = A(\Delta, \kappa, E)\delta q(t)$ (refs 22, 23), where we have neglected any noise in the system. For the parameter regime of our experiment the signal-to-noise ratio of the contribution given by the mirror's spectrum is as large as 10^7 . The dynamics of the output field quadrature is thus entirely determined by the mirror motion by means of the function $A(\Delta, \kappa, E)$. A phase-sensitive measurement of the output field quadrature δY_{out} is therefore capable of 'monitoring' the full mirror dynamics. It is particularly interesting to measure the power spectrum because $S_{Y_{\text{out}}} = \int dt' e^{i\omega t'} \langle \delta Y_{\text{out}}(t) \delta Y_{\text{out}}(t+t') \rangle = T(\Delta) S_q$, where S_q is the spectrum of the mirror motion. In other words, the quadrature power spectrum of the mirror motion S_q and of the output cavity field $S_{Y_{\text{out}}}$ are directly related by means of a transfer function $T(\Delta)$. This correspondence is at the basis of our read-out scheme. Note that the full transfer function has to take into account the sensitivity of the specific detection scheme used.

This detection strategy allows us to infer the effective temperature of the mirror's brownian motion through the study of its displacement power spectrum; that is, its frequency-dependent mean-square displacement. The power spectrum follows a lorentzian distribution centred on $\omega_0 = \sqrt{\omega_{\text{eff}}^2 - 2\gamma_{\text{eff}}^2}$ with a full-width at half-maximum (FWHM) $w_{\text{FWHM}} \approx 2\gamma_{\text{eff}}$ (for $\omega_0^2 \gg \gamma_{\text{eff}}^2$), thus proportional to the introduced damping. The area of the power spectrum, $\langle x^2 \rangle = \int_{-\infty}^{+\infty} d\omega S_q$, is proportional to the mean energy $\langle U \rangle$ of the vibrational mode and hence, by the equipartition law, to the effective temperature of the mirror, because $\langle U \rangle = m\omega_M^2 \langle x^2 \rangle = k_B T_{\text{eff}}$. The relative change in area underneath the power spectrum is therefore a direct measure of the change in effective temperature.

The system under investigation is a doubly clamped cantilever used as the end mirror of a linear optical cavity driven by an ultra-stable Nd:YAG laser (see Fig. 1). The input mirror of the cavity is attached to a piezoelectric transducer, which is fed by a control loop allowing us to lock the precise length of the cavity either at resonance or detuned (off resonance) with respect to the laser frequency. The error-signal input to the control loop is obtained with the Pound–Drever–Hall (PDH) technique²⁴. It has been shown²³ that the PDH error signal is proportional to the phase quadrature of the output field Y_{out} and hence to the mirror motion (see above). An intuitive way to view it is that the error signal is proportional to the variation of the cavity length. Above the cutoff frequency of our control loop, the fluctuations in the error signal are therefore directly related to the

thermal noise of the cantilever (the input mirror is assumed to be fixed).

We measured the PDH power spectrum for different input powers and cavity detunings. The detuning was achieved by adding an offset to the error signal. With this method, the mechanical damping can be measured directly by determining the FWHM of the resonance peak of the observed mechanical mode. To obtain the effective temperature of the mode one has to calculate the area underneath the resonance peak and account for the sensitivity of the error signal. This is done by normalizing the measured mirror amplitudes by the gradient of the PDH signal. The results are summarized in Figs 2–4.

Figure 2 shows the noise spectrum of the oscillator for two different detunings at 2 mW input laser power. The width of the peak increases and the area of the peak decreases, which is indicative of both overdamping and cooling of the mechanical mode. This behaviour is in full agreement with the theoretical model presented above. We investigate the specific variation of both mechanical damping and of self-cooling with detuning for different input laser powers of 1 and 2 mW, respectively (Figs 3 and 4). Figure 3 shows the change in width of the mechanical mode. For positive detunings, the peak is broadened from a natural width of 32 Hz to well above 800 Hz, corresponding to an extra damping of the mode. At large detuning

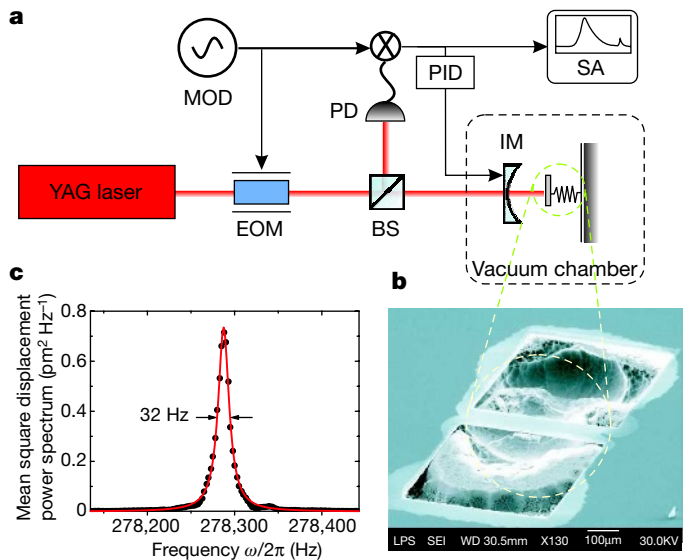


Figure 1 | Sketch of the experimental setup. **a**, A cavity was built between the cantilever and a regular concave mirror of 25 mm focal length and 99.3% reflectivity. The cavity length was slightly shorter than 25 mm so as to obtain a waist of about 20 μm at the location of the surface of the cantilever. In this configuration we measured a cavity finesse of 500. To minimize damping of the mechanical mode due to gas friction, the cavity was placed in a vacuum chamber kept at 10^{-5} mbar. The cavity was pumped with a Nd:YAG laser at 1,064 nm. The beam was phase modulated at 19 MHz (MOD) by a resonant electro-optic modulator (EOM) before it was injected into the cavity by means of the input mirror (IM). The beam reflected from the cavity was sent through a beam splitter (BS) onto a high-speed PIN photodiode (PD). After amplification of the photocurrent, its AC part was demodulated with the initial modulation frequency to obtain the PDH error signal. This error signal was then used to feed a low-frequency control loop (PID) to stabilize the cavity length by means of a piezo actuator. In addition, the error signal was fed to a spectrum analyser (SA) to record the dynamics of the mechanical mode. **b**, The cantilever was a doubly clamped free-standing Bragg mirror (520 μm long, 120 μm wide and 2.4 μm thick) that had been fabricated by using ultraviolet excimer-laser ablation in combination with a dry-etching process³⁰. The reflectivity of the Bragg mirror was 99.6% at 1,064 nm. **c**, Power spectrum of the micromirror. We isolated a mechanical mode at 280 kHz with a natural width of 32 Hz, corresponding to $Q \approx 9,000$. All measurements presented here were made on this mode. Experimental points correspond to a temperature of 300 K; solid lines are lorentzian fits to the data.

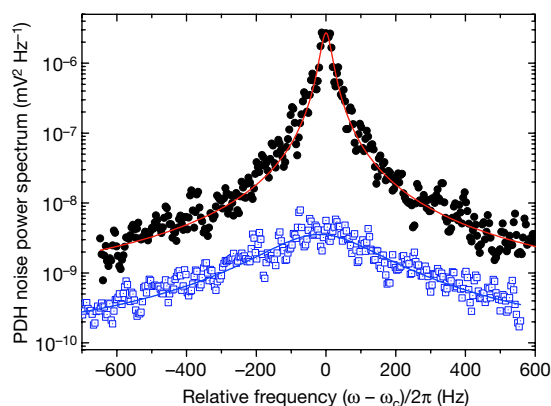


Figure 2 | Power spectrum of the mechanical mode at two different relative detuning levels Δ of the cavity for an input power of 2 mW. The data were obtained from the PDH power spectrum, which was directly proportional to the displacement power spectrum of the micromirror. Experimental points (black, $\Delta = 0$; blue, $\Delta = 0.44\kappa$) were taken with the spectrum analyser, averaged over 30 consecutive measurement runs. Solid lines are Lorentzian fits to the data. The areas obtained from the fit correspond to temperatures of 300 K (red) and 8 K (blue).

values the stability of the locking limits the precision of the measurements. For negative detuning (not shown) we observed a narrowing of the peak, associated with an amplification of the mirror motion (that is, 'negative' damping), which rapidly leads to a self-oscillation region. In Fig. 4 the same data set is used to obtain the corresponding cooling ratio from the relative change in area of the power spectrum, because the total peak area is a measure of temperature. As expected, the increase in damping is accompanied by a cooling of the mechanical mode. At large detuning, the cooling effect is slightly enhanced compared with our simple model, which might be due to the reduced contribution of thermal background of other oscillator modes. The best experimental cooling ratio in our detuning range is more than 30. Because our experiment was performed at room temperature, this corresponds to a cooling of the mode from 300 K to less than 10 K (Fig. 2).

We explicitly compare the experimental results for positive detuning with the theoretical predictions obtained if the effect is due only to radiation pressure. To do that, we have independently evaluated

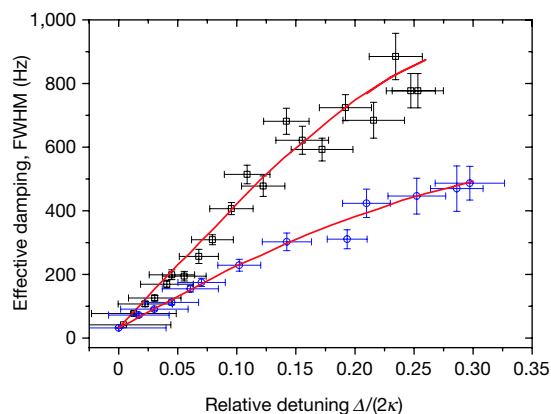


Figure 3 | Radiation-pressure-induced damping of mirror dynamics. The measured width of the mechanical mode at 278 kHz is shown at different detuning levels of the cavity and for input laser powers of 1 mW (blue points) and 2 mW (black points). The data are obtained directly from Lorentzian fits on the measured power spectra of the PDH error signal. Error bars represent absolute errors based on experimental uncertainty. Solid lines represent theoretical predictions of purely radiation-pressure effects for $F \approx 500$, $Q \approx 9,000$ and an effective mass of 9 ng. The inferred effective mass of 22 ± 4 ng indicates the presence of an additional damping force of photothermal nature (see the text).

the effective mass participating in the dynamics of the system, which leaves no free parameter for the evaluation of radiation-pressure forces and hence allows a full quantitative treatment. The effective mass can be much smaller than the total mass of the cantilever^{19,25}. For our mirror, an independent assessment both by means of spatial tomography of the vibrational mode and by means of a calibrated reference results in a value of 22 ± 4 ng at the probing point (see Supplementary Information). This results in a theoretically expected cooling that is less strong than the experimentally observed one. To get a clear, immediate figure of the 'strength' of the radiation pressure effect required to replicate the experimental data we assume a fixed effective mass and permit variation of the input power. We find that, for effective masses of 18 and 26 ng, a power respectively 2.2-fold and 3.3-fold the nominal value used in the experiment is required to match the theoretical predictions with both the observed damping and cooling (Figs 3 and 4). In other words, radiation pressure accounts for at least 30% of the observed cooling but may be as strong as 50%; that is, there is cooling by a factor between 8 and 12. We attribute the additional cooling in our setup to the presence of photothermal effects. In a similar manner to the bolometric forces reported in ref. 7, differential heating of the outer layers of the dielectric Bragg mirror can result in time-delayed changes in the cavity length, eventually introducing a retarded force that can contribute to the self-cooling mechanism. In a thin-layered medium the delayed force induced by photothermal effects can have typical time constants on the order of several tens of nanoseconds (see Supplementary Information), which is fast enough to compete with the timescale of radiation pressure effects of $1/(2\kappa)$ (about 13 ns in our experiment). The direction of the force depends on the specific material properties of the expanding layers. In our case, and in contrast with previous experiments, it assists the cooling effect of radiation pressure present for positive detuning.

The experimental data are consistent with radiation-pressure cooling assisted by photothermal effects. Residual heating of the cantilever due to absorption was not observed (see Supplementary Information). Improvements in the reflectivity of the Bragg mirror will further reduce and eventually eliminate photothermal contributions to the cooling because it will permit a higher finesse to be achieved and the optical absorption to be limited. An interesting analogy by which to understand this cooling mechanism can be found in thermodynamics. If a system (the mirror), initially at

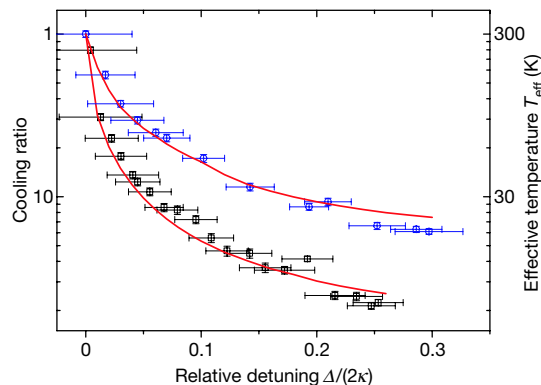


Figure 4 | Self-cooling of the mechanical resonator. The cooling ratio of the mechanical mode is shown as a function of detuning and for input laser powers of 1 mW (blue points) and 2 mW (black points). The data are obtained as the normalized area of the measured PDH power spectrum, compensated for the detuning dependent sensitivity of the PDH cavity response. Error bars represent absolute errors based on experimental uncertainty. The self-cooling effect increases for increasing laser power and detuning, in agreement with the theoretical predictions (solid lines). The right ordinate shows the inferred effective temperature of the mechanical oscillator. Radiation pressure contributes between 30% and 50% to the overall cooling, which is assisted by photothermal effects.

thermal equilibrium with a bath at ambient temperature (its environment), is strongly coupled to another bath with a very low temperature (the low-noise laser), its temperature will decrease so as to bring it to equilibrium with both baths. The current technical limitation for observing a lower temperature is the stability of the detuned locking and the base temperature from which the self-cooling starts. For example, with a cavity finesse $F = 6,000$ and an input optical power of 1 mW we expect a pure radiation-pressure cooling ratio of 1,500 for a smaller mirror oscillating at 1 MHz with an effective mass of 5 ng and $Q = 10^5$. Starting from 5 K, one should achieve cooling to 3 mK—below the base temperature of a dilution fridge. We are confident that the quantum ground state may be reachable with state-of-the-art optics and microfabrication techniques²⁶.

We have observed self-cooling of a low-mass micromirror sustained by radiation pressure. The cooling of mechanical oscillators is a key requirement for many open problems of modern physics ranging from the performance of shot-noise-limited position measurements¹ to the study of gravitational waves^{2,3} and dynamical multistability in micro-optical systems²⁷. The possibility of lowering the temperature of an oscillator to its quantum mechanical ground state paves the way to the implementation of quantum state engineering involving macroscopic systems^{21,28,29}, a closer study of the boundary between classical and quantum physics⁶ and, ultimately, the observation of non-classical correlations between macroscopic objects²⁶. In the long term it may also provide new methods of integrated quantum (mechanical) information processing.

Received 27 March; accepted 22 September 2006.

- LaHaye, M. D., Buu, O., Camarota, B. & Schwab, K. C. Approaching the quantum limit of a nanomechanical resonator. *Science* **304**, 74–77 (2004).
- Braginsky, V. & Vyatchanin, S. P. Low quantum noise tranquilizer for Fabry–Perot interferometer. *Phys. Lett. A* **293**, 228–234 (2002).
- Laser Interferometer Gravitational Wave Observatory. (<http://www.ligo.caltech.edu/>).
- Schwab, K. C. & Roukes, M. L. Putting mechanics into quantum mechanics. *Physics Today* July 36–42 (2005).
- Leggett, A. J. Testing the limits of quantum mechanics: motivation, state of play, prospects. *J. Phys. Condens. Matter* **14**, R415–R451 (2002).
- Marshall, W., Simon, C., Penrose, R. & Bouwmeester, D. Towards quantum superpositions of a mirror. *Phys. Rev. Lett.* **91**, 130401 (2003).
- Metzger, C. & Karrai, K. Cavity cooling of a microlever. *Nature* **432**, 1002–1005 (2004).
- Mertz, J. & Heidmann, A. Photon noise reduction by controlled deletion techniques. *J. Opt. Soc. Am. B* **10**, 745–752 (1993).
- Cohadon, P., Heidmann, A. & Pinard, M. Cooling of a mirror by radiation pressure. *Phys. Rev. Lett.* **83**, 3174–3177 (1999).
- Bushev, P. *et al.* Feedback cooling of a single trapped ion. *Phys. Rev. Lett.* **96**, 043003 (2006).
- Caves, C. Quantum-mechanical radiation-pressure fluctuations in an interferometer. *Phys. Rev. Lett.* **45**, 75–79 (1980).
- Braginsky, V., Strigin, S. & Vyatchanin, S. P. Parametric oscillatory instability in Fabry–Perot interferometer. *Phys. Lett. A* **287**, 331–338 (2001).
- Braginsky, V. B. *Quantum Measurements* (Cambridge Univ. Press, Cambridge, 1995).
- Tucker, R., Baney, D., Sorin, W. & Flory, C. Thermal noise and radiation pressure in MEMS Fabry–Perot tunable filters and lasers. *IEEE J. Sel. Top. Quantum Electron.* **8**, 88–97 (2002).
- Vogel, M., Mooser, C., Karrai, K. & Warburton, R. Optically tunable mechanics of microlevers. *Appl. Phys. Lett.* **83**, 1337–1339 (2003).
- Sheard, B., Gray, M., Mow-Lowry, C. & McClelland, D. Observation and characterization of an optical spring. *Phys. Rev. A* **69**, 051801 (2004).
- Dorsel, A., McCullen, J., Meystre, P., Vignes, E. & Walther, H. Optical bistability and mirror confinement induced by radiation pressure. *Phys. Rev. Lett.* **51**, 1550–1553 (1983).
- Rokhsari, H., Kippenberg, T., Carmon, T. & Vahala, K. Radiation-pressure-driven micro-mechanical oscillator. *Opt. Express* **13**, 5293–5301 (2005).
- Kippenberg, T., Rokhsari, H., Carmon, T., Scherer, A. & Vahala, K. Analysis of radiation-pressure induced mechanical oscillation of an optical microcavity. *Phys. Rev. Lett.* **95**, 033901 (2005).
- Giovanetti, V. & Vitali, D. Phase-noise measurement in a cavity with a movable mirror undergoing quantum Brownian motion. *Phys. Rev. A* **63**, 023812 (2001).
- Zhang, J., Peng, K. & Braunstein, S. L. Quantum-state transfer from light to macroscopic oscillators. *Phys. Rev. A* **68**, 013808 (2003).
- Paternoistro, M. *et al.* Reconstructing the dynamics of a movable mirror in a detuned optical cavity. *New J. Phys.* **8**, 107 (2006).
- Jacobs, K., Titttonen, I., Wiseman, H. & Schiller, S. Quantum noise in the position measurement of a cavity mirror undergoing Brownian motion. *Phys. Rev. A* **60**, 538–548 (1999).
- Black, E. D. An introduction to Pound–Drever–Hall laser frequency stabilization. *Am. J. Phys.* **69**, 79–87 (2001).
- Pinard, M., Hadjar, M. Y. & Heidmann, A. Effective mass in quantum effects of radiation pressure. *Eur. Phys. J. D* **7**, 107–116 (1999).
- Pinard, M. *et al.* Entangling movable mirrors in a double-cavity system. *Europhys. Lett.* **72**, 747–753 (2005).
- Marquardt, F., Harris, J. G. E. & Girvin, S. M. Dynamical multistability induced by radiation pressure in high-finesse micromechanical optical cavities. *Phys. Rev. Lett.* **96**, 103901 (2006).
- Bose, S., Jacobs, K. & Knight, P. L. Preparation of nonclassical states in cavities with a moving mirror. *Phys. Rev. A* **56**, 4175–4186 (1997).
- Mancini, S. & Tombesi, P. Quantum noise reduction by radiation pressure. *Phys. Rev. A* **49**, 4055–4065 (1994).
- Bäuerle, D. *Laser Processing and Chemistry* 3rd edn (Springer, Berlin, 2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Brukner, S. Gröblacher, J. Kofler, T. Jennewein, M. S. Kim, A. Vandale and D. Vitali for discussion. We acknowledge financial support by the Austrian Science Fund (FWF), by the City of Vienna, by the Austrian NANO Initiative (MNA) and by the Foundational Questions Institute (FQXi).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.A. (markus.aspelmeyer@quantum.at).

Radiation-pressure cooling and optomechanical instability of a micromirror

O. Arcizet¹, P.-F. Cohadon¹, T. Briant¹, M. Pinard¹ & A. Heidmann¹

Recent table-top optical interferometry experiments^{1,2} and advances in gravitational-wave detectors³ have demonstrated the capability of optical interferometry to detect displacements with high sensitivity. Operation at higher powers will be crucial for further sensitivity enhancement, but dynamical effects caused by radiation pressure on the interferometer mirrors must be taken into account, and the appearance of optomechanical instabilities may jeopardize the stable operation of the next generation of interferometers⁴⁻⁶. These instabilities^{7,8} are the result of a nonlinear coupling between the motion of the mirrors and the optical field, which modifies the effective dynamics of the mirror. Such 'optical spring' effects have already been demonstrated for the mechanical damping of an electromagnetic waveguide with a moving wall⁹, the resonance frequency of a specially designed flexure oscillator¹⁰, and the optomechanical instability of a silica microtoroidal resonator¹¹. Here we present an experiment where a micromechanical resonator is used as a mirror in a very high-finesse optical cavity, and its displacements are monitored with unprecedented sensitivity. By detuning the laser frequency with respect to the cavity resonance, we have observed a drastic cooling of the microresonator by intracavity radiation pressure, down to an effective temperature of 10 kelvin. For opposite detuning, efficient heating is observed, as well as a radiation-pressure-induced instability of the resonator. Further experimental progress and cryogenic operation may lead to the experimental observation of the quantum ground state of a micromechanical resonator¹²⁻¹⁴, either by passive¹⁵ or active cooling techniques¹⁶⁻¹⁸.

The resonator is placed at the end of a linear cavity, along with a conventional coupling mirror (Fig. 1a). As we are only interested in the motion at frequencies Ω close to a resonance frequency Ω_m of the resonator, the mirror dynamics can be approximated as those of a single harmonic oscillator, with resonance frequency Ω_m , mass M , damping Γ_m and mechanical susceptibility $\chi_m[\Omega]$ given by:

$$\chi_m[\Omega] = \frac{1}{M(\Omega_m^2 - \Omega^2 - i\Gamma_m\Omega)} \quad (1)$$

The resonator is submitted to a radiation-pressure force F_{rad} induced by the intracavity field. Depending on the detuning $\Psi \equiv 4\pi L/\lambda$ [2 π], where L is the cavity length and λ the laser wavelength, any small displacement x of the resonator induces a variation of the intracavity power P and of the radiation pressure (Fig. 1b). As a consequence, the spring constant $k = M\Omega_m^2$ of the resonator is balanced by the radiation-pressure force: for a positive detuning, the displacement creates a negative linear force and thereby an additional binding force, increasing the effective spring constant, whereas for a negative detuning, the force corresponds to a softening of the oscillator. Effects are null at resonance, maximum at the half-width of the optical resonance, and proportional to the incident power. These effects have already been directly observed¹⁰, as well as the bistable behaviour of

the cavity for a negative detuning¹⁹. In our experiment, however, owing to the high-finesse cavity and the high resonance frequency, such a static approach is no longer valid and one has to take into account the cavity storage time to evaluate the resonator dynamics. Additional dephasings appear, and the radiation-pressure force F_{rad} can be written in Fourier space⁶:

$$F_{\text{rad}}[\Omega] = -2 \frac{\varphi\varphi_{\text{NL}}}{\mathcal{A}} M\Omega_m^2 x[\Omega] \quad (2)$$

with

$$\mathcal{A} = (1 - i\Omega/\Omega_c)^2 + \varphi^2 \quad (3)$$

where $x[\Omega]$ is the Fourier component of the resonator displacement, $\varphi = \Psi/\gamma$ the detuning normalized to the cavity damping rate γ , and Ω_c the cavity bandwidth. φ_{NL} is a nonlinear phase-shift which characterizes the optomechanical coupling between the resonator and the light in the cavity. It corresponds to the normalized phase-shift of the cavity induced by the static recoil effect of the resonator¹⁹ and is given by:

$$\varphi_{\text{NL}} = \frac{8\pi}{\lambda\gamma c} \frac{P}{M\Omega_m^2} \quad (4)$$

At thermodynamical equilibrium, the equation of motion of the resonator is:

$$x[\Omega] = \chi_m[\Omega] (F_T[\Omega] + F_{\text{rad}}[\Omega]) \quad (5)$$

where F_T is the Langevin force responsible for the brownian motion²⁰. This equation of motion can be rewritten from equation (2) without the radiation-pressure term, but with an effective mechanical susceptibility $\chi_{\text{eff}}[\Omega]$:

$$\chi_{\text{eff}}[\Omega]^{-1} = \chi_m[\Omega]^{-1} + 2 \frac{\varphi\varphi_{\text{NL}}}{\mathcal{A}} M\Omega_m^2 \quad (6)$$

In the limit of a mechanical quality factor $Q = \Omega_m/\Gamma_m \gg 1$, $\chi_{\text{eff}}[\Omega]$ still has a lorentzian shape, but with effective resonance frequency and damping given by:

$$\Omega_{\text{eff}} = \Omega_m \left(1 + \text{Re} \frac{\varphi\varphi_{\text{NL}}}{\mathcal{A}} \right) \quad (7)$$

$$\Gamma_{\text{eff}} = \Gamma_m \left(1 - 2Q \text{Im} \frac{\varphi\varphi_{\text{NL}}}{\mathcal{A}} \right) \quad (8)$$

where $\mathcal{A}$ is now evaluated at the resonance frequency Ω_m . The Q factor in equation (8) indicates a much stronger dependence of the damping upon radiation-pressure effects as long as the imaginary part of $1/\mathcal{A}$ stays comparable to its real part, that is, for $\Omega_m \approx \Omega_c$. The radiation-pressure effects can then increase or decrease the damping of the resonator, depending on the sign of the detuning. As the Langevin force is not modified, one gets a situation somewhat similar to that obtained by cold damping¹⁶: the system still obeys the

¹Laboratoire Kastler Brossel, Université Pierre et Marie Curie, Case 74, 4 place Jussieu, F-75252 Paris Cedex 05, France.

fluctuation-dissipation theorem²⁰, but at a different effective temperature, given for small frequency shifts ($\Omega_{\text{eff}} - \Omega_m$) by:

$$\frac{T_{\text{eff}}}{T} \approx \frac{\Gamma_m}{\Gamma_{\text{eff}}} \quad (9)$$

Both radiation-pressure cooling and heating can therefore be performed, depending on the sign of the cavity detuning. A similar result has recently been demonstrated with a silicon microlever, using the photothermal force rather than radiation pressure¹⁵.

In our experiment (Fig. 2), a silicon doubly-clamped (1 mm × 1 mm × 60 μm) beam with a mirror coated upon its surface is used as a back mirror of a single-ended Fabry–Perot cavity¹⁷. We use one particular vibration mode, which has the following characteristics: $\Omega_m/2\pi \approx 814$ kHz, an effective mass $M = 190$ μg, a spring constant $k \approx 5 \times 10^6$ N m⁻¹, and a mechanical quality factor $Q = 10,000$ (Supplementary Information) in vacuum (residual pressure below 10^{-2} mbar), with the whole vacuum chamber at room temperature $T = 300$ K. Both the quality of the low-loss dielectric coating and the low roughness of the resonator have allowed us to reach an optical finesse $\mathcal{F} = \pi/\gamma = 30,000$, with a cavity bandwidth $\Omega_c/2\pi = 1.05$ MHz ($L = 2.4$ mm). Such a high finesse dramatically increases both the radiation-pressure cooling effect and the sensitivity of the phase of the reflected field to the resonator displacements. Using a highly stabilized laser source and a Pound–Drever–Hall (PDH) phase modulation scheme has indeed allowed us to reach a sensitivity of 4×10^{-19} m Hz^{-1/2} near the resonance frequency (Supplementary Information): the thermal peak of the resonator at room temperature is observed with a 50 dB signal-to-noise ratio with respect to the background thermal noise of the surrounding vibration modes.

The calibration of our set-up is performed in two steps. First, the calibration is performed when the cavity is at resonance with a frequency modulation of the laser^{2,17}. Second, we have to take into account the lowering of the sensitivity for a non-zero detuning, which stems from two origins: the lower dependence of the phase response with cavity detuning away from resonance, and the distortion of the

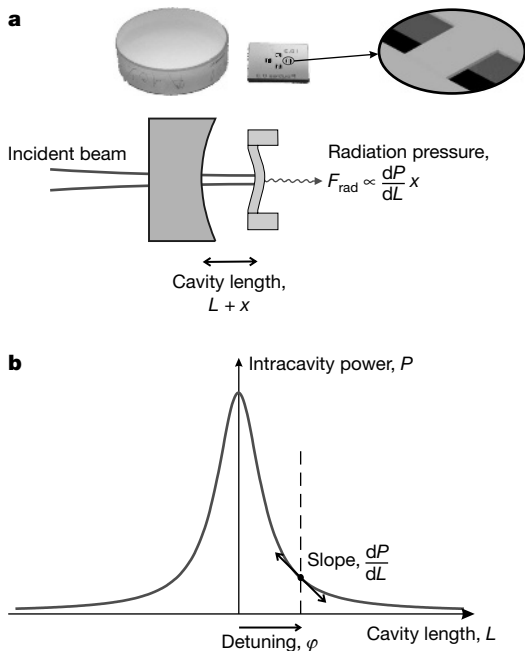


Figure 1 | Principle of radiation-pressure cooling. **a**, Layout of the optical cavity. The microresonator mirror is etched upon a 1 cm² silicon chip. The coupling mirror of the cavity is a standard low-loss silica mirror.

b, Intracavity power P as a function of the cavity length L . Around a working point defined by the normalized cavity detuning ϕ , the force F_{rad} exerted by the intracavity field upon the microresonator is proportional to the slope of the Airy peak and to the displacement x of the resonator.

PDH signal for large detunings. For that purpose, we drive the resonator into motion (with an amplitude of $\sim 10^{-13}$ m) with a modulated electrostatic force at 814 kHz and monitor the displacement signal for different detunings, and an incident power (50 μW) low enough to ensure that radiation pressure has no effect. The observed modulation has been used to calibrate every spectrum observed.

Figure 3 shows the thermal noise spectra obtained for negative detunings and a 5 mW incident beam (Fig. 3a), and for positive detunings and a 2.5 mW incident beam (Fig. 3b). In both cases, the black curve corresponds to the resonant situation $\phi = 0$. This thermal noise spectrum (r.m.s. amplitude $\sim 2.9 \times 10^{-14}$ m, driven by a 8×10^{-25} N² Hz⁻¹ Langevin force spectral density) has been used to infer the value of the effective mass (190 μg) used throughout this Letter. This value is in good agreement with a finite-element method (FEM) computation, which yields¹⁷ a value of 130 μg. The discrepancy is accounted for by the overall accuracy of both the FEM computation and the fabrication of the microresonator, by the imperfect overlap of the vibration mode with a non-centred optical beam, and by the coupling of the mode with the ones of the wafer the resonator is engraved onto.

Both cooling and heating are evident on the noise spectra of Fig. 3. For a negative detuning, the spectra are both widened and drastically decreased at their resonance frequencies. The decrease of the area of the curves, which is directly related to the effective temperature by the equipartition theorem, is a strong indication of the temperature reduction. The situation is opposite for positive detunings, where the spectra are strongly narrowed and increased at resonance. The absence of any spurious noise in the spectra is due to both the high sensitivity of our experiment and the stable operation of the cavity. Note also in both cases the clear frequency shift ($\Omega_{\text{eff}} - \Omega_m$) of the resonances.

Our experimental results can be further tested by looking at the dependence of both the damping ratio $\Gamma_{\text{eff}}/\Gamma_m$ and the frequency shift $(\Omega_{\text{eff}} - \Omega_m)/2\pi$ with respect to the detuning ϕ . Figure 4a presents our experimental results, for five optical powers from 0.5 mW to 3.2 mW, along with the theoretical fits deduced from equations (7) and (8). Both are in very good agreement. The inset shows the intracavity power at resonance P , the single free parameter derived from the preceding fits, as a function of the incident power P_{in} . It exhibits the expected linear dependence, with $P/P_{\text{in}} = 2,970$ in agreement at the 1% level with the value deduced from cavity and insertion losses. This clearly shows that the observed cooling effect is solely due to

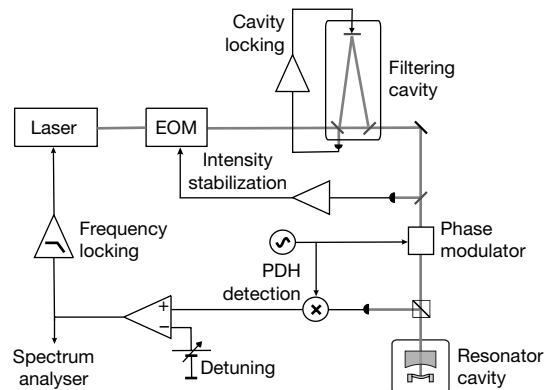


Figure 2 | Experimental set-up used to monitor and to cool the micromechanical oscillator. A Nd:YAG laser ($\lambda = 1,064$ nm) is intensity-stabilized at low frequency with an electro-optic modulator (EOM) and spatially filtered before entering the microresonator cavity. The displacement signal is extracted by means of a Pound–Drever–Hall (PDH) phase modulation scheme using a resonant electro-optic phase modulator. The low-frequency part of the signal is used to lock the laser frequency around the cavity resonance, with a preset detuning given by a d.c. offset. The high-frequency part is used to monitor the resonator displacements.

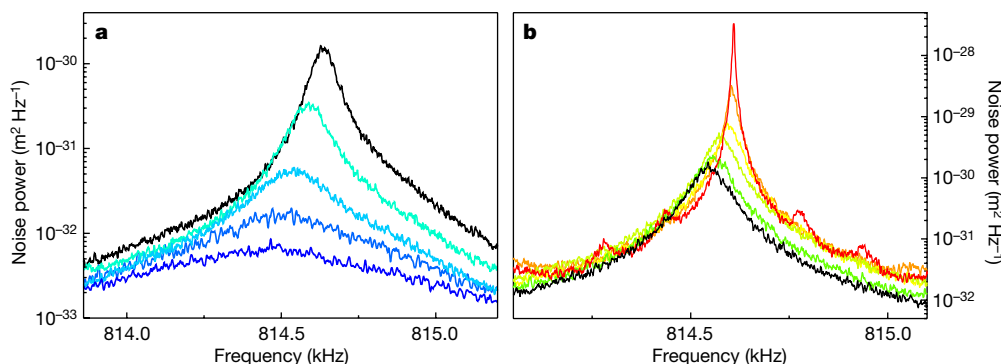


Figure 3 | Thermal noise spectra, normalized as microresonator displacements. Black curves correspond to $\varphi = 0$. **a**, Curves light to dark blue are obtained for negative detunings $\varphi = -0.1, -0.25, -0.4$ and -0.6 , respectively, and for an incident power of 5 mW. **b**, Curves green to red are

obtained for positive detunings $\varphi = 0.03, 0.06, 0.09, 0.11$ and 0.13 , respectively, and for an incident power of 2.5 mW. The cooling and heating are evident through the area reduction or increase of these spectra. Note the related drift of the resonance frequency.

radiation pressure. Indeed, with the estimated mirror absorption (~ 3 p.p.m.), thermoelastic effects are expected to be 10^6 times lower. Bolometric effects¹⁵ are difficult to estimate, but both the low absorption and the large thickness of the resonator are in favour of a negligible effect. We have also performed the same experiment for other modes of the resonator, with the same excellent agreement with theory. In particular, a mode with a resonance frequency $\Omega_m = 2\pi \times 2.824$ MHz larger than the cavity bandwidth Ω_c exhibits a frequency shift ($\Omega_{\text{eff}} - \Omega_m$) with an opposite sign, as expected from equations (3) and (7) for not-too-large detunings φ .

Figure 4b presents a colour-chart of the dependence of the effective damping Γ_{eff} in the detuning/intracavity power plane. Experimental

series of points were taken for fixed stabilized incident powers, following the Airy curves in the plane. The colour code ranges from dark blue (for large Γ_{eff} and low T_{eff}) to red (for low Γ_{eff} and large T_{eff}). The green points (corresponding to $\Gamma_{\text{eff}} \approx \Gamma_m$) are located at the resonance, at large detunings (where the slope of the Airy peak vanishes) and at low intracavity power (where radiation-pressure effects are weaker). The highest cooling (dark blue) is obtained near a negative detuning $\varphi \approx -0.6$ and for a high intracavity power P , whereas the largest heating effect (red) is obtained for a positive detuning. Grey curves correspond to equal effective dampings Γ_{eff} and the black one to $\Gamma_{\text{eff}} = 0$. Above this line, the resonator becomes unstable and starts to oscillate at its effective resonance frequency, as

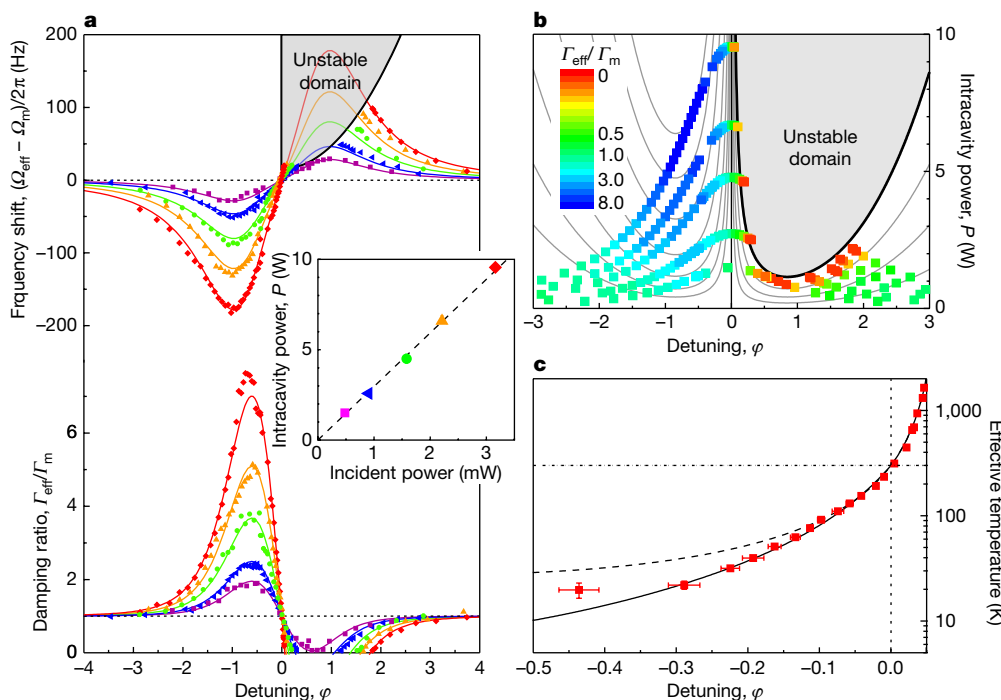


Figure 4 | Evolution of the cavity cooling and heating effects with respect to the detuning φ . **a**, Frequency shift $(\Omega_{\text{eff}} - \Omega_m)/2\pi$ (top) and damping ratio $\Gamma_{\text{eff}}/\Gamma_m$ (bottom), for five values of incident power: 0.5 mW (purple), 0.9 mW (blue), 1.6 mW (green), 2.2 mW (yellow) and 3.2 mW (red). Points are experimental results and full lines are fits obtained from equations (4), (7) and (8), by adjusting the intracavity power at resonance for each curve. Inset, evolution of the adjusted intracavity power P with the incident power. The dashed line is the linear dependence expected from cavity and insertion losses, with no adjustable parameter. The shaded area in the upper curve shows the instability zone where Γ_{eff} vanishes. **b**, Colour-chart of the evolution of the damping ratio $\Gamma_{\text{eff}}/\Gamma_m$ in the detuning/intracavity power

plane $\{\varphi, P\}$, for the same measurements as in **a**. The value is colour-coded from dark blue (large damping, low temperature) to red (low damping, high temperature). Grey curves are equal effective damping loci. Note the green points (unity damping ratio) at the resonance $\varphi = 0$ and the red points in the vicinity of the instability region (shaded area). **c**, Evolution of the effective temperature with the cavity detuning, for a 3.2 mW incident beam. Squares, experimental points with error bars (s.e.m.). The dotted line is the reference level at $T = 300$ K ($\varphi = 0$), the dashed line the fit with the single oscillator model of equation (9), and the full line the fit with a model taking into account the out-of-resonance tails of other modes of the microresonator.

has already been demonstrated with a microtoroidal resonator¹¹. We have observed a similar effect, as shown on Fig. 4a and b, where the instability region is displayed with no adjustable parameter. Also note that the secondary peaks on the red spectrum of Fig. 3b (more than 50 dB below the resonance spectrum level) are due to the lower stability of the cavity operation close to the instability region. This is to our knowledge the first experimental demonstration of such a radiation-pressure instability in an open optical cavity.

Figure 4c displays the variations of the effective temperature with the detuning ϕ . For a fixed input power of 3.2 mW, T_{eff} ranges from 20 K at an optimum detuning $\phi \approx -0.45$, to 2,000 K for a positive detuning, very close to the instability region. The dependence with ϕ is well accounted for by our single oscillator theoretical model (equation (9)) for positive and negative but not too large detunings (dashed line). For larger negative detunings, the noise spectrum is so low that the thermal noise background due to other vibration modes of the resonator is no longer negligible: a slight modification of our model taking this background into account depicts well the characteristics of our experimental results (solid line). The residual discrepancy at large detuning (leftmost point in Fig. 4c) cannot be related to a residual heating process since effective temperatures as low as 10 K have been obtained with a 12 mW incident beam.

Our experimental set-up is to our knowledge the first high-sensitivity optical displacement sensor to show a direct effect of radiation pressure, a ubiquitous—though extremely weak—back-action effect that every optical experiment will eventually (once all technical problems are resolved) be sensitive to, once limited only by quantum noise. Further enhancement and low-temperature operation of such a system may lead to the experimental demonstration of the quantum ground state of a mechanical resonator^{12–15,17,18}; radiation-pressure cooling is a fundamental process that involves nothing but a mechanical oscillator in interaction with one cavity mode and its quantum fluctuations, yielding a reliable estimation of the cooling limit. Furthermore, our knowledge of the other modes' dynamical behaviour¹⁷ appears as a major issue on the road to the quantum ground state, as their out-of-resonance tails will no longer be negligible for extreme cooling ratios. Other possible novel effects in quantum optics include the experimental demonstration of the standard quantum limit^{21,22}, radiation-pressure-induced squeezing of light²³, quantum non-demolition measurements²⁴ or non-classical states of the resonator motion^{25,26}.

Received 19 May; accepted 8 September 2006.

1. Rugar, D., Budakian, R., Mamin, H. J. & Chui, B. W. Single spin detection by magnetic resonance force microscopy. *Nature* **430**, 329–332 (2004).
2. Hadjar, Y., Cohadon, P.-F., Aminoff, C. G., Pinard, M. & Heidmann, A. High-sensitivity measurement of mechanical Brownian motion. *Europhys. Lett.* **47**, 545–551 (1999).
3. Abbott, B. *et al.* Upper limits on a stochastic background of gravitational waves. *Phys. Rev. Lett.* **95**, 221101 (2005).
4. Meers, B. Recycling in laser-interferometric gravitational-wave detectors. *Phys. Rev. D* **38**, 2317–2326 (1988).
5. Heinzel, G. *et al.* Experimental demonstration of a suspended dual recycling interferometer for gravitational wave detection. *Phys. Rev. Lett.* **81**, 5493–5496 (1998).

6. Arcizet, O., Briant, T., Heidmann, A. & Pinard, M. Beating quantum limits in an optomechanical sensor by cavity detuning. *Phys. Rev. A* **73**, 033819 (2006).
7. Pai, A., Dhurandar, S., Hello, P. & Vinet, J.-Y. Radiation pressure induced instabilities in laser interferometric detectors of gravitational waves. *Eur. Phys. J. D* **8**, 333–346 (2000).
8. Zhao, C., Ju, L., Degallaix, J. & Blair, D. Parametric instabilities and their control in advanced interferometer gravitational-wave detectors. *Phys. Rev. Lett.* **94**, 121102 (2004).
9. Braginsky, V. B., Manukin, A. B. & Tikhonov, M. Yu. Investigation of dissipative ponderomotive effects of electromagnetic radiation. *Sov. Phys. JETP* **31**, 829–830 (1970).
10. Sheard, B. S., Gray, M. B., Mow-Lowry, C. M., McClelland, D. E. & Whitcomb, S. E. Observation and characterization of an optical spring. *Phys. Rev. A* **69**, 051801 (2004).
11. Kippenberg, T. *et al.* Analysis of radiation-pressure induced mechanical oscillation of an optical microcavity. *Phys. Rev. Lett.* **95**, 033901 (2005).
12. Huang, X. M. H., Zorman, C. A., Mehregany, M. & Roukes, M. L. Nanoelectromechanical systems: Nanodevice motion at microwave frequencies. *Nature* **421**, 496 (2003).
13. Knobel, R. G. & Cleland, A. N. Nanometre-scale displacement sensing using a single electron transistor. *Nature* **424**, 291–293 (2003).
14. LaHaye, M. D., Buu, O., Camarota, B. & Schwab, K. C. Approaching the quantum limit of a nanomechanical resonator. *Science* **304**, 74–77 (2004).
15. Hohberger Metzger, C. & Karrai, K. Cavity cooling of a microlever. *Nature* **432**, 1002–1005 (2004).
16. Cohadon, P.-F., Heidmann, A. & Pinard, M. Cooling of a mirror by radiation pressure. *Phys. Rev. Lett.* **83**, 3174–3177 (1999).
17. Arcizet, O. *et al.* High-sensitivity optical monitoring of a micro-mechanical resonator with a quantum-limited optomechanical sensor. *Phys. Rev. Lett.* **97**, 133601 (2006).
18. Cohadon, P.-F., Arcizet, O., Briant, T., Heidmann, A. & Pinard, M. Optical monitoring and cooling of a micro-mechanical oscillator to the quantum limit. *Proc. SPIE* **5846**, 124–134 (2005).
19. Dorsel, A., McCullen, J. D., Meystre, P., Vignes, E. & Walther, H. Optical bistability and mirror confinement induced by radiation pressure. *Phys. Rev. Lett.* **51**, 1550–1553 (1983).
20. Kubo, R. The fluctuation-dissipation theorem. *Rep. Prog. Phys.* **29**, 255–284 (1966).
21. Caves, C. M. Quantum-mechanical noise in an interferometer. *Phys. Rev. D* **23**, 1693–1708 (1981).
22. Jaekel, M. T. & Reynaud, S. Quantum limits in interferometric measurements. *Europhys. Lett.* **13**, 301–306 (1990).
23. Fabre, C. *et al.* Quantum-noise reduction using a cavity with a movable mirror. *Phys. Rev. A* **49**, 1337–1343 (1994).
24. Heidmann, A., Hadjar, Y. & Pinard, M. Quantum nondemolition measurement by optomechanical coupling. *Appl. Phys. B* **64**, 173–180 (1997).
25. Bose, S., Jacobs, K. & Knight, P. L. Preparation of nonclassical states in cavities with a moving mirror. *Phys. Rev. A* **56**, 4175–4186 (1997).
26. Pinard, M. *et al.* Entangling movable mirrors in a double-cavity system. *Europhys. Lett.* **72**, 747–753 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to L. Rousseau for the fabrication of the microresonator and to J.-M. Mackowski and his group for the optical coating of the resonator. This work was partially funded by EGO (collaboration convention for a study of quantum noises in gravitational-wave interferometers). Laboratoire Kastler Brossel is a research unit of Ecole Normale Supérieure and Université Paris 6, associated with CNRS.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.-F. C. (cohaddon@spectro.jussieu.fr).

Sub-kelvin optical cooling of a micromechanical resonator

Dustin Kleckner¹ & Dirk Bouwmeester¹

Micromechanical resonators, when cooled down to near their ground state, can be used to explore quantum effects such as superposition and entanglement at a macroscopic scale^{1–3}. Previously, it has been proposed to use electronic feedback to cool a high frequency (10 MHz) resonator to near its ground state⁴. In other work, a low frequency resonator was cooled from room temperature to 18 K by passive optical feedback⁵. Additionally, active optical feedback of atomic force microscope cantilevers has been used to modify their response characteristics⁶, and cooling to approximately 2 K has been measured⁷. Here we demonstrate active optical feedback cooling to 135 ± 15 mK of a micromechanical resonator integrated with a high-quality optical resonator. Additionally, we show that the scheme should be applicable at cryogenic base temperatures, allowing cooling to near the ground state that is required for quantum experiments—near 100 nK for a kHz oscillator.

Using a laser tuned to the resonance fringe of a high finesse optical cavity, it is possible to observe very small fluctuations in the length of the cavity due to brownian motion of one or both of the end mirrors. We have developed an optical cavity with one rigid large mirror, 6 mm in diameter and with a 25 mm radius of curvature, and one tiny plane mirror, 30 μ m in diameter, attached to a commercial atomic force microscope cantilever of dimensions $450 \times 50 \times 2$ μ m with a fundamental resonance of 12.5 kHz (Fig. 1b). An optical finesse of 2,100 and a mechanical quality factor of 137,000 have been achieved with the system⁸. The motion of the tiny mirror/cantilever is monitored by measuring the transmission of the cavity at a frequency on the side of an optical resonance peak. To do this, we use about 1 mW from a 780 nm tunable diode laser which is locked to the resonance fringe using the integrated signal from a photo-multiplier tube which monitors the light transmitted through the cavity (Fig. 1a). The time derivative of this signal is proportional to the velocity of the cantilever tip and is used to modulate the amplitude of a second, 980 nm, diode laser focused on the cantilever less than 100 μ m away from the tiny mirror. The radiation pressure exerted by this feedback laser counteracts the motion of the mirror and effectively provides cooling of the fundamental mode.

The effective feedback gain can be varied over several orders of magnitude by sending the feedback laser through a variable neutral density filter. The average power in the feedback beam when it reaches the cantilever is of the order of 1 mW at the highest gain settings and proportionally lower otherwise. The mean modulation depth of the feedback beam varies from nearly 100% to less than 5% as gain is increased. The vibration spectrum of the cantilever as a function of gain is shown in Fig. 2. The r.m.s. thermal amplitude of the cantilever without feedback is 1.2 ± 0.1 Å. From this value, one can calculate that the spring constant of the cantilever is 0.15 ± 0.01 N m^{–1}, in agreement with the manufacturer-specified

range, and the effective mass of the cantilever fundamental mode is $(2.4 \pm 0.2) \times 10^{-11}$ kg.

To determine the effective gain of the feedback loop and the temperature of the fundamental mode, we fit a lorentzian plus a constant background to the vibration spectrum of the cantilever for each value of feedback gain. The temperature is determined from the area under the lorentzian without the background, while the gain is determined by the width of the resonance. The linewidth provides a good measure of gain because it is directly determined by the damping rate whereas the cantilever amplitude may be affected by other sources of noise in the feedback loop. Cooling is observed over more than three orders of magnitude. The lowest temperature we are able to measure is 135 ± 15 mK, or a cantilever r.m.s. amplitude of 0.023 ± 0.002 Å, with a gain (the ratio of feedback to mechanical damping) of $g = 2,490 \pm 90$ (Fig. 2b). The lowest trace in Fig. 2b, indicating an even lower temperature, cannot be reliably fitted owing to the laser noise floor. Since the optical finesse is not the current limiting factor, we operate the opto-mechanical system at a finesse of only 200,

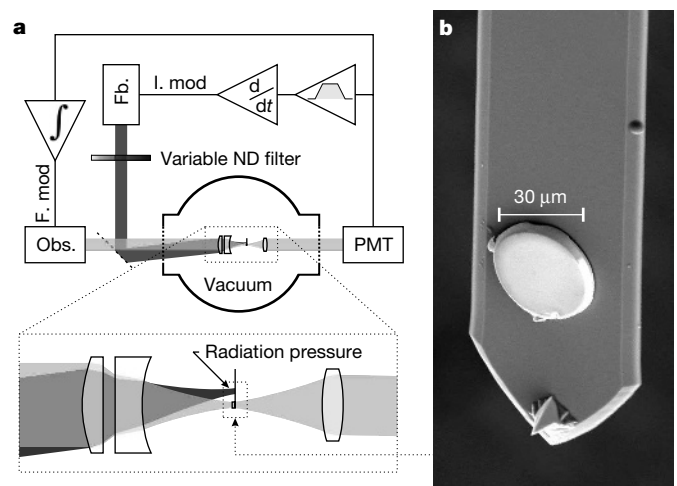


Figure 1 | The experimental system. **a**, Diagram of the feedback mechanism: a 780 nm observation laser (Obs.) is frequency locked to the optical cavity (shown magnified at bottom) with an integrating circuit (via the laser frequency modulation input, f. mod), using the signal from a photomultiplier tube (PMT). This signal is also sent through a 1.25 kHz bandpass filter at 12.5 kHz and a derivative circuit (d/dt) to provide an intensity-modulating signal (I. mod.) for the 980 nm feedback laser (Fb.). The feedback laser is attenuated with a variable neutral density (ND) filter to adjust the gain of the feedback. The feedback force is exerted on the cantilever via this laser's radiation pressure. **b**, Scanning electron microscope image of the tip of the cantilever with attached mirror.

¹Department of Physics, University of California, Santa Barbara, California 93106, USA.

produced by slight cavity misalignment, which makes the system less sensitive to transient vibrations.

The amplitude of the mirror motion can be calculated in the presence of feedback by assuming that the Langevin force—the effective thermal force that maintains brownian motion—remains constant while the mechanical susceptibility of the mirror is reduced by the dissipation due to the radiation feedback pressure. It suffices to consider only the fundamental mode of the mirror motion, represented by a damped harmonic oscillator. In this approximation, the power spectrum of the mirror's motion in the presence of feedback becomes⁹:

$$S_x^{\text{fb}}[\Omega] = \frac{2\Gamma_0 k_B T_0}{M} \frac{1}{(\omega^2 - \Omega^2)^2 + (1+g)^2 \Gamma_0^2 \Omega^2} \quad (1)$$

where Ω is the observation frequency, ω is the resonator frequency, Γ_0 is the mechanical damping factor, M is the effective mass of the resonator mode, k_B is Boltzmann's constant, T_0 is the bulk temperature of the resonator and g is the gain. $g=0$ corresponds to the vibration spectrum in the absence of feedback. The motion of the oscillator in the presence of feedback is the same as that of an oscillator with lower temperature and a higher damping constant:

$$T_{\text{fb}} = (1+g)^{-1} T_0 \quad (2)$$

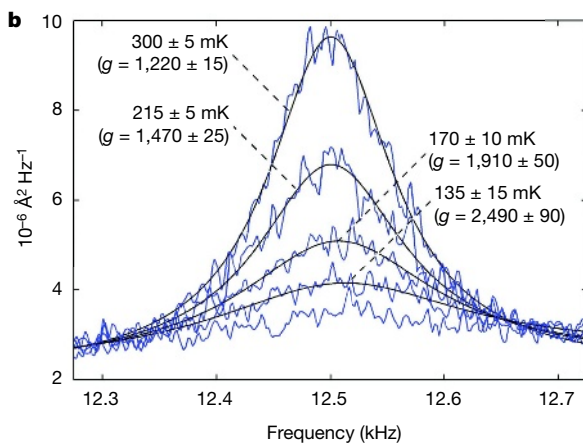
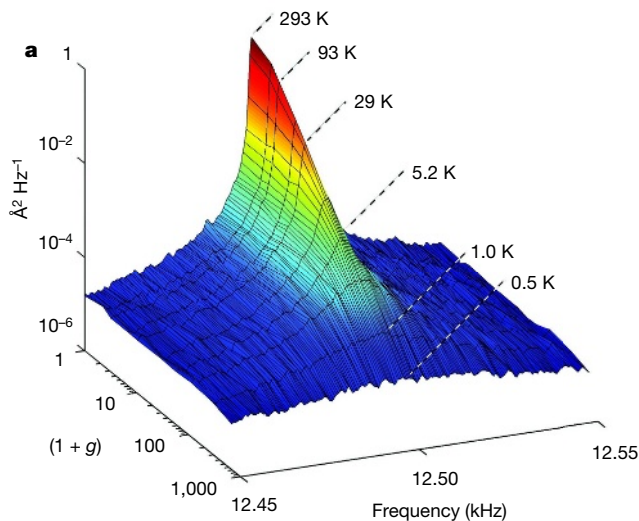


Figure 2 | Single-sided thermal vibration spectrum of the cantilever as it is cooled. g is the dimensionless gain factor, which is the ratio of feedback to mechanical damping. **a**, Spectrum at low to moderate gains. **b**, Spectrum near the background noise level for large gains. The blue curves correspond to experimental data, and the black curves to fits of a gaussian function plus a background. The lowest trace cannot be reliably fitted.

$$\Gamma_{\text{fb}} = (1+g)\Gamma_0 \quad (3)$$

The optical feedback scheme, when analysed in terms of noiseless classical light fields, can be seen as a virtual viscous force, which unlike a real viscous force creates dissipation without introducing fluctuations. As discussed below, the cooling temperature as demonstrated here is limited by laser frequency fluctuations. Ultimately, optical cooling should be limited by the balance of residual heating and quantum noise in the observation and feedback laser signals.

For a signal-to-noise ratio of one in spectral density at the peak of the mechanical resonance, the temperature of the cantilever would be (as can be derived from equation (1)):

$$T_{\text{min}} \cong \sqrt{\frac{T_0 M \omega^3 S_{\text{noise}}}{2k_B Q}} \quad (4)$$

where S_{noise} is the equivalent position noise in the interferometer measurement and $Q = \omega/\Gamma_0$ is the mechanical quality factor. For higher values of gain, the feedback signal is mostly noise and lower temperatures can not be conclusively demonstrated. For our experiment, the equivalent noise level is $\sqrt{S_{\text{noise}}} \approx 10^{-3} \text{ Å Hz}^{-1/2}$. This corresponds to the expected noise due to the frequency fluctuations of a free running tunable laser diode, which are of order $10^3 \text{ Hz Hz}^{-1/2}$ at the resonance frequency of 12.5 kHz (ref. 10). With the system in vacuum at pressures of 10^{-6} mbar, so as to maximize the mechanical quality factor of the cantilever, this noise level corresponds to a minimum temperature of the order of 100 mK, in good agreement with the experimental data.

An alternative approach to study the cooling is to analyse the temporal response of the system by gating the signal to the feedback laser. The characteristic time constant for the system to reach equilibrium after the cooling is turned on is given by:

$$\tau_{\text{fb}} = \Gamma_{\text{fb}}^{-1} = (1+g)^{-1} \Gamma_0^{-1} \quad (5)$$

To observe this behaviour, we monitor the cantilever over many 10 s periods during each of which the cooling is on for 3 s. Data for cooling to 1.8 ± 0.2 , 4.0 ± 0.2 and 6.4 ± 0.1 K and returning to thermal equilibrium are shown in Fig. 3. The cooling times are measured to be 9.0 ± 0.5 , 19 ± 1 and 27 ± 1 ms, respectively. The reheating time is found to be indistinguishable for all three gains with an average of $\tau_0 = 1.30 \pm 0.05$ s. This is in agreement with the linewidth of the cantilever measured without feedback, $\Gamma_0 = 680 \pm 50$ mHz. In

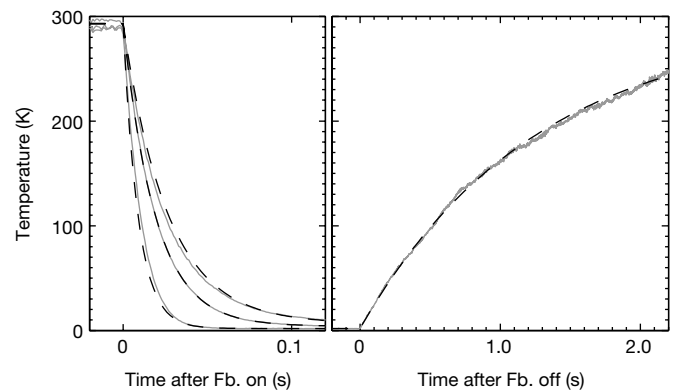


Figure 3 | Temporal response of the cantilever to cooling pulses. The temperature is determined by calculating the total vibrational amplitude of the cantilever between 12 and 13 kHz in 1 ms bins and subtracting the background. Each data set is the average of 1,000 samples. The three sets in the left panel correspond to cooling to 6.4, 4.0 and 1.8 K (solid lines, top to bottom). Heating is shown (right panel) for only one data set (1.8 K), as all three are nearly coincident. The dashed lines are fits to exponential decays, used to determine the cooled temperature and the cooling and reheating times. Fb. refers to the feedback system.

accordance with theory, the ratio of the reheating to the cooling times, τ_0/τ_{fb} , and the corresponding ratio of the spectral linewidths from the earlier measurements, Γ_{fb}/Γ_0 , are found to be the same as the cooling factor, T_0/T_{fb} , within statistical uncertainties.

In experiments where optical feedback is used on cantilevers with non-uniform composition, radiation pressure is typically overwhelmed by the photothermal force, which is an effective force due to thermally induced bending^{5,6}. Although this is not the case for single-crystal silicon cantilevers, the addition of a tiny mirror on the tip of our cantilever should produce a weak photothermal force. This force can be distinguished from radiation pressure by its dependence on the intensity modulation frequency of the feedback laser. Whereas radiation pressure is independent of modulation frequency, the photothermal force is not, because it has a characteristic response time, τ , related to the thermal relaxation time of the cantilever. A simple model for the frequency dependence of the photothermal force, $F_{pt}(\Omega)$, gives:

$$F_{pt}(\Omega) \cong \int_0^\infty \frac{F_{pt}(0)}{\tau} e^{-t/\tau} e^{-i\Omega t} dt = \frac{F_{pt}(0)}{1+i\Omega\tau} \quad (6)$$

where $e^{-i\Omega t}$ corresponds to the input power modulation, and $e^{-t/\tau}$ is due to the thermal relaxation. This is consistent with the frequency dependence of the photothermal force as described in previous work⁶. To test for the presence of photothermal force in our resonator, the feedback laser was modulated at a range of frequencies from 100 Hz to 20 kHz and the mechanical response of the cantilever was measured as before (Fig. 4). The power in the feedback laser reflected from the cantilever was determined to have a mean of 2.7 ± 0.5 mW and a modulation amplitude of 1.0 ± 0.2 mW, independent of the modulation frequency. This results in a radiation pressure force of $F_{rad} = 2P_{mod}/c = 6.7 \pm 1.3$ pN (where P_{mod} is the amplitude of the power modulation and c is the speed of light) at the modulation frequency.

If the driving frequency is sufficiently far from the cantilever resonance, the mechanical damping constant can be ignored and the amplitude of the cantilever's motion should be of the form:

$$A(\Omega) = \left| \frac{\frac{A_{pt}}{1+i\Omega\tau} + A_{rad}}{1 - (\Omega/\omega)^2} \right| \quad (7)$$

where Ω is the driving frequency, ω is the resonance frequency, τ is the photothermal characteristic time, and A_{rad} and A_{pt} are the magnitudes of the motion due to the radiation pressure and photothermal force alone, at zero frequency. The term in the denominator is due to mechanical amplification by the cantilever resonance. This equation fits well to the measured response (Fig. 4), resulting in

$A_{rad} = 0.470 \pm 0.005 \text{ \AA}$, $A_{pt} = -6.3 \pm 0.2 \text{ \AA}$ and $\tau = 30 \pm 2$ ms. At frequencies greater than 5 kHz, radiation pressure is observed to be the dominant force mechanism, whereas the photothermal force is relevant only at lower frequencies.

Assuming the constant force background described by A_{rad} is entirely due to radiation pressure, one can calculate the spring constant of the cantilever at the position where the feedback laser is focused to be $k = F_{rad}/A_{rad} = 0.14 \pm 0.03 \text{ N m}^{-1}$, in agreement with the value for the spring constant obtained earlier. Near the fundamental resonance of the cantilever, the radiation pressure is calculated to be almost 5 times larger than the photothermal force. Additionally, the two forces should be nearly 90° out of phase at this frequency, given that the time constant of the photothermal force is found to be 30 ± 2 ms. Thus the radiation pressure is responsible for almost all of the demonstrated feedback cooling; in the absence of photothermal force, the total feedback force would be reduced by less than 3%.

When optical cooling is active, the cantilever's motion is strongly damped, making it undesirable for many types of measurements. In some cases this problem can be overcome with a stroboscopic cooling scheme, where measurements are only made in the periods when the cooling is off. In addition to being of direct importance for the aforementioned massive superposition experiment, this scheme has already been theoretically shown to be useful for high sensitivity measurements of position and weak impulse forces¹¹. Because the cooling is faster than the heating by a factor $(1+g)$, a low temperature can be maintained even when the cooling is off the majority of the time. However, maintaining low temperatures requires that the measurement window be short; if it is, for example, one oscillation period long, the temperature of the oscillator will have increased by $\Delta T \approx 2\pi T_0/Q$ by the end of each measurement window, meaning that cooling past this point results in marginal improvement.

We now evaluate the potential for reaching even lower temperatures for the purpose of studying quantum effects in similar systems. Reference 3 proposes an experiment: putting a mechanical oscillator in a quantum superposition of vibrating and not-vibrating by interaction with the light pressure of a single photon in an optical cavity of which one end mirror is attached to the oscillator. Appropriate for such a scheme would be a 250- μm -long silicon cantilever with a 20- μm -diameter dielectric mirror on the tip and a resonance frequency of 1 kHz. Because of the constraints of environmentally induced decoherence^{12,13}, the bulk temperature must be less than $T_{EID} = Q\hbar\omega/k_B = 8 \text{ mK}$ for the cantilever to remain coherent over one period, given $Q \approx 150,000$. This temperature is achievable by conventional means; nuclear adiabatic demagnetization of PrNi_5 (ref. 14) could be employed as the final, vibration-free, cooling stage, as it is able to be started from temperatures previously demonstrated

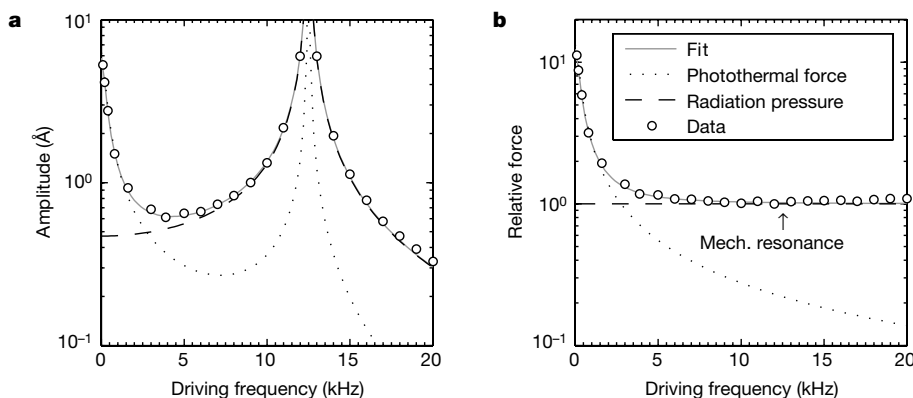


Figure 4 | Response of the cantilever to an external intensity-modulated laser. **a**, The amplitude of the cantilever's motion at the driving frequency. **b**, The force on the cantilever, calculated by dividing the amplitude by the mechanical amplification of the cantilever. In both graphs the magnitude of

the contributions (ignoring phase differences) of the photothermal force and radiation pressure are shown as dashed and dotted lines, respectively. The slight deviation of the fit from the data at higher frequencies is due to higher-order flexural modes.

for vibration-isolated cold stages (~ 100 mK)^{15,16}. The observation period for a massive superposition experiment is one oscillation long, thus the maximum useful cooling factor is $Q/2\pi \approx 25,000$ as discussed above. This corresponds with a temperature of 300 nK or a mean oscillator quantum number of only 2π .

It has been shown theoretically that optical feedback still works in the quantum regime, allowing cooling to the ground state¹⁷. Experimentally, cooling to the quantum regime requires the capability to accurately monitor the position of the cantilever without introducing significant heating. With an optical finesse $F = 5 \times 10^5$, which should be technologically achievable⁸, an observation beam power of 1 aW, or about 5,000 photons per second, is enough to reduce shot noise to the appropriate level. Assuming the thermal conductivity of the cantilever is reduced to the one-dimensional quantum limit, the cantilever's thermal resistivity will be roughly 30 mK aW^{-1} (ref. 18). The heating from the feedback laser can be reduced by use of a sufficiently long wavelength laser so that absorption is negligible; this is not possible for the readout beam, which must be resonant with the optical cavity. Thus as long as the observation laser has relatively low absorption in the cantilever/mirror, it should not significantly affect the bulk temperature. This implies that cooling a kHz oscillator to near its ground state should be possible, drastically simplifying the experimental requirements to observe quantum phenomena in this system.

We have demonstrated active laser feedback cooling of a micro-mechanical oscillator, using only radiation pressure, from room temperature to 135 mK. Furthermore, we have shown that this cooling method could be used in addition to traditional cryogenics to reach much lower temperatures, even near the ground state of a kHz oscillator. This in turn would significantly aid the realization of proposals to create and investigate massive quantum superpositions.

Received 5 June; accepted 24 August 2006.

1. Bose, S., Jacobs, K. & Knight, P. L. Scheme to probe the decoherence of a macroscopic object. *Phys. Rev. A* **59**, 3204–3210 (1999).
2. Mancini, S., Vitali, D., Giovannetti, V. & Tombesi, P. Stationary entanglement between macroscopic mechanical oscillators. *Eur. Phys. J. D* **22**, 417–422 (2003).

3. Marshall, W., Simon, C., Penrose, R. & Bouwmeester, D. Towards quantum superpositions of a mirror. *Phys. Rev. Lett.* **91**, 130401 (2003).
4. Hopkins, A., Jacobs, K., Habib, S. & Schwab, K. Feedback cooling of a nanomechanical resonator. *Phys. Rev. B* **68**, 235238 (2003).
5. Metzger, C. H. & Karrai, K. Cavity cooling of a microlever. *Nature* **432**, 1002–1005 (2004).
6. Mertz, J., Marti, O. & Mlynek, K. Regulation of a microcantilever response by force feedback. *Appl. Phys. Lett.* **62**, 2344–2346 (1993).
7. Bruland, K. J., Garbini, J. L., Dougherty, W. M. & Sidles, J. A. Optimal control of force microscope cantilevers. II. Magnetic coupling implementation. *J. Appl. Phys.* **80**, 1959–1964 (1996).
8. Kleckner, D. *et al.* High finesse opto-mechanical cavity with a movable thirty-micron-size mirror. *Phys. Rev. Lett.* **96**, 173901 (2006).
9. Cohadon, P. F., Heidmann, A. & Pinard, M. Cooling of a mirror by radiation pressure. *Phys. Rev. Lett.* **83**, 3174–3177 (1999).
10. Turner, L. D., Weber, W. P., Hawthorn, C. J. & Scholten, R. E. Frequency noise characterisation of narrow linewidth diode lasers. *Opt. Commun.* **201**, 391–397 (2002).
11. Vitali, B. D., Mancini, S., Ribichini, L. & Tombesi, P. Mirror quiescence and high-sensitivity position measurements with feedback. *Phys. Rev. A* **65**, 063803 (2002).
12. Joos, E. *et al.* *Decoherence and the Appearance of a Classical World in Quantum Theory* 2nd edn (Springer, New York, 2003).
13. Zurek, W. H. Decoherence and the transition from quantum to classical. *Phys. Today* **44**, 36–44 (1991).
14. Parpia, J. M. *et al.* Optimization procedure for the cooling of liquid ^3He by adiabatic demagnetization of praseodymium nickel. *Rev. Sci. Instrum.* **56**, 437–443 (1985).
15. Moussy, N., Courtois, A. & Pannetier, B. A very low temperature scanning tunnelling microscope for the local spectroscopy of mesoscopic structures. *Rev. Sci. Instrum.* **71**, 128–131 (2001).
16. Mamin, H. J. & Rugar, D. Sub-attoneutron force detection at millikelvin temperatures. *Appl. Phys. Lett.* **79**, 3358–3360 (2001).
17. Courty, J. M., Heidman, A. & Pinard, M. Quantum limits of cold damping with optomechanical coupling. *Eur. Phys. J. D* **17**, 399–408 (2001).
18. Schwab, K., Henriksen, E. A., Worlock, J. M. & Roukes, M. L. Measurement of the quantum of thermal conductance. *Nature* **404**, 974–977 (2000).

Acknowledgements This work was supported by the National Science Foundation. We thank M. de Dood, H. Eisenberg, S. Hastings, W. Irvine, A. Kahl, G. Khoury, W. Marshall and C. Simon for their contributions at earlier stages of this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.K. (dkleckner@physics.ucsb.edu).

Complex zeolite structure solved by combining powder diffraction and electron microscopy

Fabian Gramm¹, Christian Baerlocher¹, Lynne B. McCusker¹, Stewart J. Warrender², Paul A. Wright², Bada Han³, Suk Bong Hong³, Zheng Liu^{4,5}, Tetsu Ohsuna^{4†} & Osamu Terasaki⁴

Many industrially important materials, ranging from ceramics to catalysts to pharmaceuticals, are polycrystalline and cannot be grown as single crystals. This means that non-conventional methods of structure analysis must be applied to obtain the structural information that is fundamental to the understanding of the properties of these materials. Electron microscopy might appear to be a natural approach, but only relatively simple structures have been solved by this route. Powder diffraction is another obvious option, but the overlap of reflections with similar diffraction angles causes an ambiguity in the relative intensities of those reflections. Various ways of overcoming or circumventing this problem have been developed^{1,2}, and several of these involve incorporating chemical information into the structure determination process^{3–7}. For complex zeolite structures, the FOCUS algorithm^{8,9} has proved to be effective. Because it operates in both real and reciprocal space, phase information obtained from high-resolution transmission electron microscopy images can be incorporated directly into this algorithm in a simple way. Here we show that by doing so, the complexity limit can be extended much further. The power of this approach has been demonstrated with the solution of the structure of the zeolite TNU-9 ($[\text{H}_{9.3}][\text{Al}_{9.3}\text{Si}_{182.7}\text{O}_{384}]$; ref. 10) with 24 topologically distinct (Si,Al) atoms and 52 such O atoms. For comparison, ITQ-22 (ref. 11), the most complex zeolite known to date, has 16 topologically distinct (Si,Ge) atoms.

To compensate, at least in part, for the loss of information caused by the collapse of the three-dimensional single-crystal diffraction data onto the single axis (2θ or d -spacing) of a powder diffraction experiment, chemical information (for example, bond distances, angles, connectivity) can be used to augment the diffraction data in a structure determination program. In the case of zeolites, the connectivity of the atoms is not known, but typical Si–O bond distances, and typical O–Si–O and Si–O–Si bond angles, are known. Furthermore, it is known that a zeolite framework structure is three-dimensional and that all Si atoms are four-connected to neighbouring Si atoms via O bridges. This is the information used in the program FOCUS⁸, which was written with the aim of solving complex zeolite structures from powder diffraction data. The algorithm combines an exhaustive framework search with Fourier recycling, and has indeed proven to be effective for structures with up to 12 T-atoms (tetrahedral framework atoms) in the asymmetric unit⁹.

The strategy involved is relatively simple. Reflection intensities are extracted from the powder diffraction pattern as well as possible (usually the intensities of overlapping reflections are simply equipartitioned), and random phases, consistent with the symmetry, are assigned to the structure factors of these reflections. An electron density map is generated from this information (random phases

essentially produce a random structure), and then interpreted automatically. From this (partial) model, a new set of phases is calculated, and if these are different from those used to generate the previous map, a new one using the new phases and the original intensities is generated and interpreted. This Fourier recycling loop is repeated until either the phases converge or a maximum number of cycles is reached. Then a new set of random phases is generated and the whole process is repeated. Each time an electron density map is produced, a search for a complete framework structure is conducted. If one is found, it is classified and written to a file. The framework structure found most frequently will usually be the correct one. One of the key aspects of the FOCUS algorithm is that it operates in both real (electron density map interpretation) and reciprocal (phases of the reflections) space. This means that additional information can be included in either realm in a straightforward manner.

Electron microscopy techniques can also be used to study very small crystallites, and indeed the structure of the zeolite SSZ-48 was solved directly from electron diffraction intensities¹². By combining high-resolution transmission electron microscopy (HRTEM) images, electron diffraction patterns and Patterson techniques, the structure of a BEC-type¹³ material overgrown on another zeolite¹⁴ has been determined. However, zeolite structure solution using electron crystallography in this way is relatively rare¹⁵. On the other hand, it is often easy to detect where the channels of a zeolite structure lie in an HRTEM image (for example, see Fig. 1A). Such information can be incorporated into the FOCUS algorithm in real space in the form of a structure envelope¹⁶. This is a periodic nodal surface¹⁷ that divides the unit cell into regions of high and low electron density, and can be generated from just a few (1–10) strong reflections with low indices. By restricting the framework search to just the positive side of such a surface, the FOCUS algorithm can be made more efficient¹⁸. However, to generate the envelope, the phases for the reflections are required. Whereas phase information is lacking in a conventional X-ray diffraction experiment, whether single-crystal or powder, it can be obtained from HRTEM images. The envelope information can also be invoked in reciprocal space in FOCUS by imposing the phases of these few reflections on each starting phase set (phases for all other reflections are generated randomly). All phases are then allowed to change during the Fourier recycling procedure, and this can be important should some of the imposed phases be incorrect.

It is not always possible to obtain good HRTEM images in all three projections, so the phase information may only be two-dimensional (that is, only one zone). However, a single projection usually contains phase information on a large number of reflections, not just the few that are required for a structure envelope. Reasoning that this information could also be used to advantage in FOCUS, we

¹Laboratory of Crystallography, ETH Zurich, 8093 Zurich, Switzerland. ²School of Chemistry, University of St Andrews, St Andrews, Fife KY16 9ST, UK. ³Division of Applied Chemistry and Biotechnology, Hanbat National University, Taejeon 305-719, Korea. ⁴Arrhenius Laboratory, Stockholm University, 10691 Stockholm, Sweden. ⁵National Institute of Advanced Industrial Science & Technology (AIST), Tsukuba, Ibaraki 3058565, Japan. [†]Present address: Department of Applied Chemistry, Waseda University, Ohkubo-3, Shinjuku-ku, Tokyo 169-8555, Japan.

performed a series of tests on experimental and simulated data. The results showed clearly that the phase information facilitated structure solution. For ITQ-22 (ref. 11), with 16 T-atoms in the asymmetric unit, the computing time required to solve the structure could be reduced from ~ 31 days for a single correct solution (no phase information used) to 44 hours for two correct solutions (31 phases input). When 124 phases were used, the correct solution was found 2,248 times in just 12 hours.

The high-silica zeolite TNU-9 (Taejon National University number 9, Si/Al = 19.6) was synthesized using 1,4-bis(*N*-methylpyrrolidinium)butane and Na^+ ions as the structure-directing agents, according to procedures described elsewhere¹⁰. Initially, its powder diffraction pattern could not be indexed even though extremely high-resolution synchrotron data had been obtained on a rehydrated calcined sample at the European Synchrotron Radiation Facility in Grenoble⁹. Only with the help of selected-area electron diffraction data (Fig. 1) did it become possible to derive a unit cell and a possible space group ($C2/m$; $a = 28.2219 \text{ \AA}$, $b = 20.0123 \text{ \AA}$, $c = 19.4926 \text{ \AA}$, $\beta = 92.33^\circ$), and to assign indices to the peaks in the powder pattern. Even then, it was not clear that the indexing was correct, because

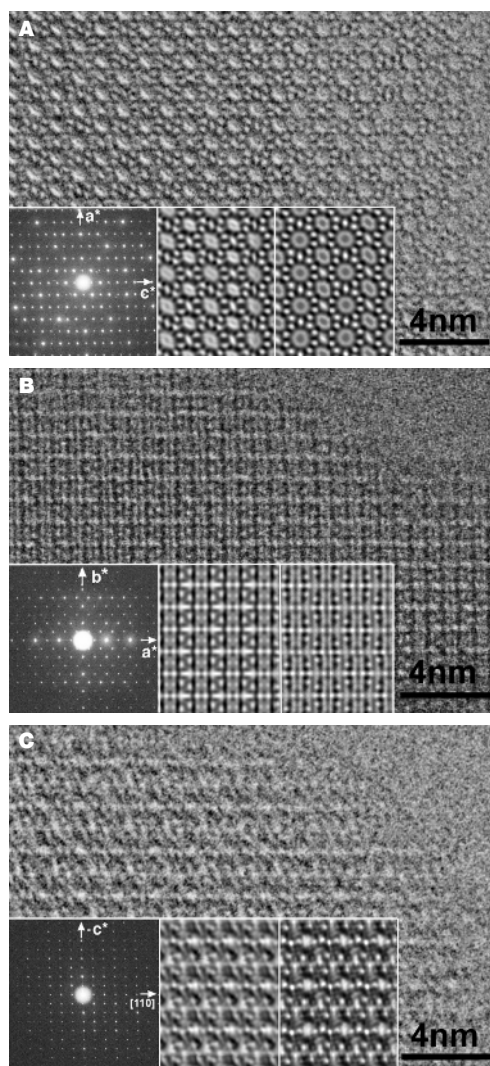


Figure 1 | High-resolution transmission electron microscopy images taken along different zone axes of TNU-9. A, [010]; B, [001]; and C, [110]. Shown as insets are the corresponding selected-area electron diffraction pattern (left), the CTF-corrected and symmetry-averaged image from which the phases were calculated (middle) and the computer simulation from the structural data (right). Sample thicknesses for the simulations were 4, 4 and 14 nm, respectively.

there were unindexed lines present (subsequent syntheses showed that these were indeed from an impurity).

Extremely high quality HRTEM images were recorded for three different projections of the TNU-9 structure using a 300 kV JEM-3010 microscope. These were corrected for the contrast transfer function (CTF) and symmetry-averaged using the CRISP software package²⁰ (Fig. 1). The projection down [010] is tantalizingly similar to that of ZSM-5²¹ (MFI framework type¹³), but the others are obviously different. Although the images appear to have close to atomic resolution, none of the many attempts to solve the structure by simple model building on the basis of these images was successful.

From the size of the asymmetric unit, it was estimated that TNU-9 would have 24 symmetrically independent Si atoms. The solution of a structure of this complexity from powder diffraction data alone appeared to be beyond the limits of current methods. Not only is the symmetry low and the unit cell very large (approximately double that of ZSM-5) with a correspondingly large number of reflections, but the lattice parameters are closely related to one another ($a \approx \sqrt{2}b$ and $b \approx c$), so symmetrically unrelated reflections have similar d -spacings and therefore similar diffraction angles. Even with the very-high-resolution data, 3,154 of the 3,705 reflections ($d_{\min} = 1.17 \text{ \AA}$) are within $0.3 \times \text{FWHM}$ of another reflection, forming 637 groups of overlapping reflections. A set of reflection intensities was extracted using a Pawley-type algorithm²² (expected to be less sensitive to the impurity peaks than a LeBail-type²³ extraction). To optimize the partitioning of overlapping reflections as far as possible, the extracted intensities were repartitioned using the fast iterative Patterson squaring method²⁴. Then a number of FOCUS runs with different sets of input parameters were performed. As anticipated, however, no solution was obtained. Even with a structure envelope (generated using 21 phases from the three HRTEM images), there was no success (both real and reciprocal space implementations were tried).

Finally, in an attempt to utilize all of the information that could be extracted from the HRTEM images, all of the phases from all three zones (258 reflections, including the weak reflections) and

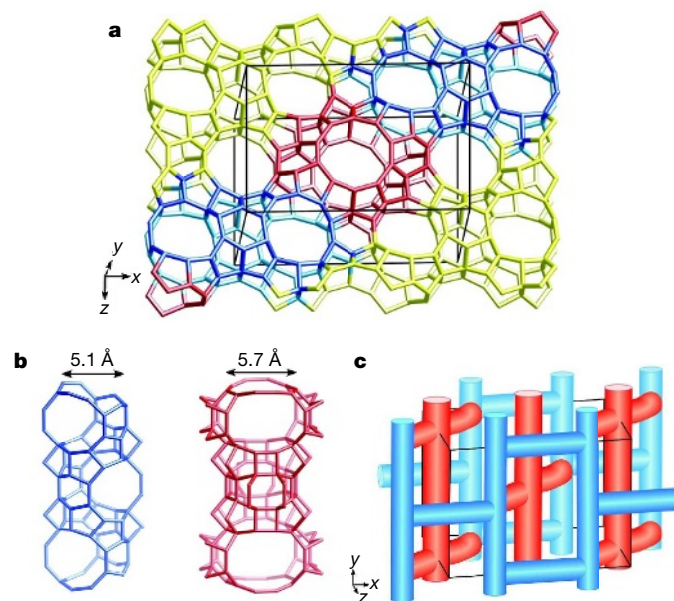


Figure 2 | The framework structure of TNU-9 (O atoms have been omitted for clarity). **a**, A layer of the structure showing the Si-atom connectivity, and the zig-zag course of the 10-ring channel running perpendicular to the b axis (highlighted in blue and red). **b**, The two types (blue and red) of 10-ring channels running parallel to the b axis with their approximate free diameters. The colours match those in **a**. **c**, A schematic drawing of the three-dimensional channel system using the same colour coding as in **a** and **b**.

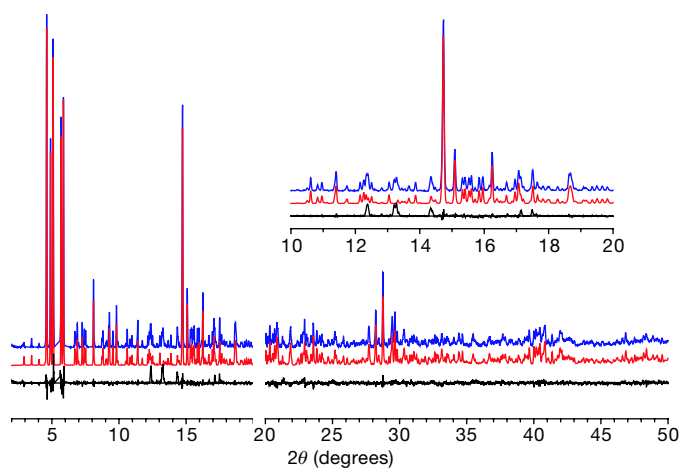


Figure 3 | Observed (blue), calculated (red) and difference (black) profiles for the Rietveld refinement of TNU-9. To show more detail, the 20–50° 2 θ range has been scaled up by a factor of four. Inset, an expanded view of the 10–20° 2 θ range, where most of the impurity peaks appear.

the strongest reflections from the powder diffraction data (representing 65% of the total structure factor amplitudes) were used as input to FOCUS. After 16 days of computing time (Mac G5 Xserve), the structure was found. The positions of all 24 Si atoms were located, and then it was a simple matter of adding O bridges between the Si atoms to complete the structure (Fig. 2). The structure was verified with a full Rietveld refinement using powder diffraction data collected on a calcined (not rehydrated) sample at the Daresbury synchrotron facility. The R values converged to $R_F = 0.057$ and $R_{wp} = 0.156$. Crystallographic data are provided in Supplementary Information. The major differences in the profile fit (Fig. 3) can be attributed to the impurity phase.

With 24 crystallographically and topologically distinct Si atoms, the framework structure of TNU-9 is by far the most complex zeolite framework known to date. The projection of TNU-9 down the b axis is very similar to that of ZSM-5, but the connectivity in the third direction is more complex. The resulting three-dimensional 10-ring channel system (Fig. 2c) is reminiscent of that of ZSM-5, but it does have some unique features. A full account of the synthesis, characterization and catalytic properties of TNU-9 will be given elsewhere.

To our knowledge, this is the first time that powder diffraction and HRTEM data have been used simultaneously in a structure determination procedure. The power of this combination is readily apparent from the complexity of the TNU-9 structure, with 72 degrees of freedom (24 Si atoms $\times$ 3 atomic coordinates) in the initial structure solution and 76 atoms in the final structure. To extend this method to include other (non-zeolite) materials (for example, pharmaceuticals, organometallics, ceramics) and thereby realize its full potential, a more general structure solution algorithm that includes both a reciprocal- and a real-space component is currently being developed.

Received 12 May; accepted 30 August 2006.

- David, W. I. F., Shankland, K., McCusker, L. B. & Baerlocher, C. (eds) *Structure Determination from Powder Diffraction Data* (Oxford Univ. Press, Oxford, 2002).
- Baerlocher, Ch. & McCusker, L. B. (eds) *Structure determination from powder diffraction data*. *Z. Kristallogr.* **219** (Spec. Iss.), 782–901 (2004).

- Harris, K. D. M., Habershon, S., Cheung, E. Y. & Johnston, R. L. Developments in genetic algorithm techniques for structure solution from powder diffraction data. *Z. Kristallogr.* **219**, 838–846 (2004).
- Favre-Nicolin, V. & Cerny, R. A better FOX: using flexible modelling and maximum likelihood to improve direct-space *ab initio* structure determination from powder diffraction. *Z. Kristallogr.* **219**, 847–856 (2004).
- Burton, A. W. Structure solution of zeolites from powder diffraction data. *Z. Kristallogr.* **219**, 866–880 (2004).
- Florence, A. J. et al. Solving molecular crystal structures from laboratory X-ray powder diffraction data with DASH: the state of the art and challenges. *J. Appl. Crystallogr.* **38**, 249–259 (2005).
- Brodski, V., Peschar, R. & Schenk, H. Organa – a program package for structure determination from powder diffraction data by direct-space methods. *J. Appl. Crystallogr.* **38**, 688–693 (2005).
- Grosse-Kunstleve, R. W., McCusker, L. B. & Baerlocher, C. Powder diffraction data and crystal chemical information combined in an automated structure determination procedure for zeolites. *J. Appl. Crystallogr.* **30**, 985–995 (1997).
- Grosse-Kunstleve, R. W., McCusker, L. B. & Baerlocher, C. Zeolite structure determination from powder diffraction data: applications of the FOCUS method. *J. Appl. Crystallogr.* **32**, 536–542 (1999).
- Hong, S. B. et al. Synthesis, structure solution, characterization, and catalytic properties of TNU-10: a high-silica zeolite with the STI topology. *J. Am. Chem. Soc.* **126**, 5817–5826 (2004).
- Corma, A., Rey, F., Valencia, S., Jorda, J. L. & Rius, J. A zeolite with interconnected 8–10- and 12-ring pores and its unique catalytic selectivity. *Nature Mater.* **2**, 493–497 (2003).
- Wagner, P. et al. Electron diffraction structure solution of a nano-crystalline zeolite at atomic resolution. *J. Phys. Chem. B* **103**, 8245–8250 (1999).
- Baerlocher, C., Meier, W. M. & Olson, D. H. *Atlas of Zeolite Framework Types* (Elsevier, Amsterdam, 2001).
- Ohsumi, T., Liu, Z., Terasaki, O., Hiraga, K. & Cambor, M. Framework determination of a polytype of zeolite beta by using electron crystallography. *J. Phys. Chem. B* **106**, 5673–5678 (2002).
- Dorset, D. L., Roth, W. J. & Gilmore, C. J. Electron crystallography of zeolites – the MWV family as a test of direct 3D structure determination. *Acta Crystallogr. A* **61**, 516–527 (2005).
- Brenner, S., McCusker, L. B. & Baerlocher, C. Using a structure envelope to facilitate structure solution from powder diffraction data. *J. Appl. Crystallogr.* **30**, 1167–1172 (1997).
- von Schnering, H. G. & Nesper, R. Nodal surfaces of Fourier series: fundamental invariants of structured matter. *Z. Phys. B* **83**, 407–412 (1991).
- Brenner, S., McCusker, L. B. & Baerlocher, C. The application of structure envelopes in structure determination from powder diffraction data. *J. Appl. Crystallogr.* **35**, 243–252 (2002).
- Fitch, A. N. The high resolution powder diffraction beam line at ESRF. *J. Res. NIST* **109**, 133–142 (2004).
- Hovmöller, S. CRISP: crystallographic image processing on a personal computer. *Ultramicroscopy* **41**, 121–135 (1992).
- Kokotailo, G. T., Lawton, S. L., Olson, D. H. & Meier, W. M. Structure of synthetic zeolite ZSM-5. *Nature* **272**, 437–438 (1978).
- Pawley, G. S. Unit-cell refinement from powder diffraction scans. *J. Appl. Crystallogr.* **14**, 357–361 (1981).
- Le Bail, A., Duroy, H. & Fourquet, J. L. Ab-initio structure determination of LiSbWO_4 by X-ray powder diffraction. *Mater. Res. Bull.* **23**, 447–452 (1988).
- Estermann, M. A. & Gramlich, V. Improved treatment of severely or exactly overlapping Bragg reflections for the application of direct methods to powder data. *J. Appl. Crystallogr.* **26**, 396–404 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the beamline scientists at SRS, Daresbury, and A. Fitch at the ESRF, Grenoble, for their assistance with the powder diffraction measurements. Funding from the Swiss National Science Foundation (L.B.M., C.B., F.G.), the Swedish Science Research Council and the Japan Science and Technology Agency (T.O., O.T.), and the Korea Science and Engineering Foundation (S.B.H.) is acknowledged.

Author Contributions F.G., C.B. and L.B.M. performed the data analysis and structure solution; S.J.W. and P.A.W. collected the synchrotron powder diffraction data and coordinated the project; Z.L., T.O. and O.T. obtained the HRTEM images; and B.H. and S.B.H. synthesized and characterised TNU-9.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.B.M. (mccusker@mat.ethz.ch).

LETTERS

¹⁰Be evidence for the Matuyama–Brunhes geomagnetic reversal in the EPICA Dome C ice core

G. M. Raisbeck¹, F. Yiou¹, O. Cattani² & J. Jouzel²

An ice core drilled at Dome C, Antarctica, is the oldest ice core so far retrieved¹. On the basis of ice flow modelling and a comparison between the deuterium signal in the ice with climate records from marine sediment cores, the ice at a depth of 3,190 m in the Dome C core is believed to have been deposited around 800,000 years ago², offering a rare opportunity to study climatic and environmental conditions over this time period. However, an independent determination of this age is important because the deuterium profile below a depth of 3,190 m depth does not show the expected correlation with the marine record². Here we present evidence for enhanced ¹⁰Be deposition in the ice at 3,160–3,170 m, which we interpret as a result of the low dipole field strength during the Matuyama–Brunhes geomagnetic reversal, which occurred about 780,000 years ago. If correct, this provides a crucial tie point between ice cores, marine cores and a radiometric timescale.

Inversions of the Earth's geomagnetic dipole represent a well-established geochronological framework. The most recent of these inversions, referred to as the Matuyama–Brunhes (M–B) transition, has been dated radiometrically to have occurred 776 ± 12 (taking into account decay constant and calibration uncertainties) kyr ago³. This transition has also been dated astrochronologically to be 778 ± 2 kyr ago⁴. However, as emphasized in ref. 5, in addition to the assumptions inherent in such a technique, it involves several sources of uncertainty (variable sedimentation rates, natural remanent magnetization (NRM) acquisition depth, and so on) that are difficult to quantify. Several authors have also reported evidence for a 'precursor' event ~15 kyr before the M–B boundary (refs 6–8 and references therein). Thus whether the M–B transition is a fairly long and complex event³ or two shorter events separated by ~15 kyr seems uncertain at present^{7,8}.

During a geomagnetic reversal, the dipole field strength is believed to decrease by about an order of magnitude⁹. During this time, galactic cosmic rays can more easily penetrate into the Earth's atmosphere and thus increase the production of cosmogenic isotopes, such as ¹⁰Be (ref. 10). This effect has indeed been observed in marine sediments¹¹. The quantitative increase of ¹⁰Be expected in polar ice cores is uncertain. Because geomagnetic field lines are vertical at the geomagnetic poles, changes in the dipole strength will have essentially no effect on cosmogenic production rates at latitudes $>60^\circ$. Thus, the only way that geomagnetic changes could have been recorded in cosmogenic isotopes deposited in polar ice is if a significant fraction of them were produced at lower latitudes. On the basis of nuclear weapons fallout patterns, Lal and Peters¹⁰ have argued that such a contribution is not important. At least some interpretations of ¹⁰Be deposition in Antarctica seem to support this position¹². On the other hand, the strong enhancement of ¹⁰Be observed in both Arctic and Antarctic ice ~40 kyr ago^{13–15}, which is now generally agreed to be coincident with the Laschamp geomagnetic excursion, seems to

contrast with this prediction. A Global Climate Model calculation suggests that, for modest reductions in the dipole intensity under present climatic conditions, the increase in ¹⁰Be deposition at high latitudes would be about 80% of the global increase¹⁶.

Thus, although the expected enhancement of ¹⁰Be in Antarctica due to the M–B reversal is uncertain, it certainly seemed worthwhile to look for it in the EPICA Dome C core. In addition to contributing to the ice core chronology, such an identification could potentially give information on the intensity structure of a reversal at unprecedented time resolution. Volcanic rocks can provide good absolute paleointensities, but they are by nature not continuous. Marine sediments are continuous, but their records are limited by such things as overprinting, smoothing, NRM acquisition depth and bioturbation.

A series of 11-cm-long samples, each representing 100–200 yr, was used for the ¹⁰Be measurements (see Methods section). The results, together with a deuterium profile², which is a proxy for climate, are shown in Fig. 1. As we have shown previously, because ¹⁰Be deposition on the Antarctic plateau is largely by dry deposition, concentrations are inversely correlated with the snow accumulation rate¹⁷. Thus a more appropriate parameter to use when studying production variations, especially over different climatic periods, is the ¹⁰Be flux. We have done this in Fig. 1 by combining the measured concentrations with accumulation rates estimated as described in ref. 1. For the flux curve of Fig. 1 we have also corrected for the radioactive decay of ¹⁰Be, using an extended version of the EDC2 timescale¹ (see Methods). This correction varies only from 1.37 to 1.45 over the interval of interest, so the conclusions are not very sensitive to dating uncertainties.

The most striking features in Fig. 1 are the numerous sharp spikes in ¹⁰Be. These spikes usually occur in only a single 11-cm sample, but can have amplitudes as much as an order of magnitude larger than adjacent samples. We have never observed such events in younger ice at Dome C or in our previous studies, at similar time resolution, in ice at Vostok, over the past ~300 kyr (refs 13, 14 and unpublished work). At the present time, the origin of these spikes is unknown. Measurements of duplicate samples at the same depths as seven of these spikes have shown: (1) cases in which no spike was observed; (2) cases where the spike was confirmed; and (3) cases where the spike was confirmed but it occurred at the same depth as spikes in other species (dust, calcium, sulphate, and so on). Thus, it appears unlikely that these spikes are caused by production variations, but rather represent some sort of local concentration effect. However, their presence makes it very difficult to evaluate the production rate trends in the ¹⁰Be profile, which is what interests us here. While investigation into the origin of these spikes continues, our immediate goal is to find an objective way of eliminating their influence on the longer-term trends.

The most obvious way of doing this would be simply to remove the spikes. However, this requires a subjective decision for what level a

¹Centre de Spectrométrie Nucléaire et de Spectrométrie de Masse, IN2P3-CNRS-Université de Paris-Sud, Bât. 108, 91405 Orsay, France. ²Laboratoire des Sciences du Climat et de l'Environnement/Institut Pierre Simon Laplace, CEA-CNRS-UVSQ, CE Saclay, 91191 Gif-sur-Yvette, France.

given result should be considered a spike. Such a procedure would also bias the profile if the spikes were simply concentrating ^{10}Be from adjacent depths. In the latter case, the most appropriate procedure would be to smooth the data, for example by running means or a low-pass filter. However, because the spikes are so large, we have found that these procedures give unsatisfactory results even when smoothing over several thousand years. This seems to confirm our visual impression that the spikes are in fact not caused by simply concentrating the ^{10}Be from adjacent depths. Thus, if the concentration effect mentioned above is the explanation for the spikes, the concentration must be predominantly in the horizontal direction. We speculate that this may be associated with the strong orientation of the large ice crystals at these depths¹⁸.

One well-known way of eliminating the effect of 'outliers' on the average of a distribution of results is to use medians rather than means. We have therefore adopted a procedure where we take the median of the five results from each 55-cm 'bag' sample. These results are also shown in Fig. 1. As can be seen, the profile is much more regular, but does show a significant peak between 3,160 and 3,170 m. This corresponds to 775–786 kyr in the EDC2 timescale, which is shown on the upper abscissa of Fig. 1. Thus, the peak is consistent with both the age and duration expected for the M–B transition. To show this more clearly, we plot in Fig. 2 the ^{10}Be median flux curve together with an estimate of the palaeomagnetic dipole intensity constructed by stacking a number of records from marine sediment

cores¹⁹. Each of the curves is plotted on its independent timescale. Because ^{10}Be production is inversely proportional to the dipole field, the intensity record is inverted. As can be seen, there is a remarkable similarity, not only in the age and duration of the peak at the M–B reversal, but even in some of the structure of the curves. For example, it is tempting to associate the shorter period of ^{10}Be enhancement at 3,185 m (~800 kyr ago) with a similar structure in the palaeomagnetic curve, which itself might correspond to the 'precursor' event mentioned above^{6–8}. There are no other periods of sustained high ^{10}Be flux, but there are several high points grouped around 3,100 m (~700 kyr ago), which again correspond to a period of apparently low dipole intensity.

If the above interpretation is correct, it provides strong independent support for the EDC2 timescale. The peak we identify as the M–B reversal occurs towards the end of what has been assumed in ref. 2 to be the equivalent of marine isotope stage 19. This is consistent with some observations in marine sediments^{20,21} but not with others, which observe it near the beginning of this stage²², or even in stage 20 (ref. 23). The authors of this last study have attributed the lower-than-expected stratigraphic position to a delayed acquisition of the magnetic signal in the sediment. The required 'acquisition depth' to account for the assumed discrepancy between their observed and expected depth is ~1 m, which is considerably larger than most estimates of this parameter. This illustrates an advantage of the ice core compared to marine cores for placing the M–B boundary in the correct stratigraphic position with respect to climatic records.

As mentioned earlier, the expected amplitude of ^{10}Be enhancement in polar ice owing to a reduced dipole depends on the fraction deposited that was produced at latitudes lower than ~60°. This parameter is uncertain even for present climatic conditions, and there is

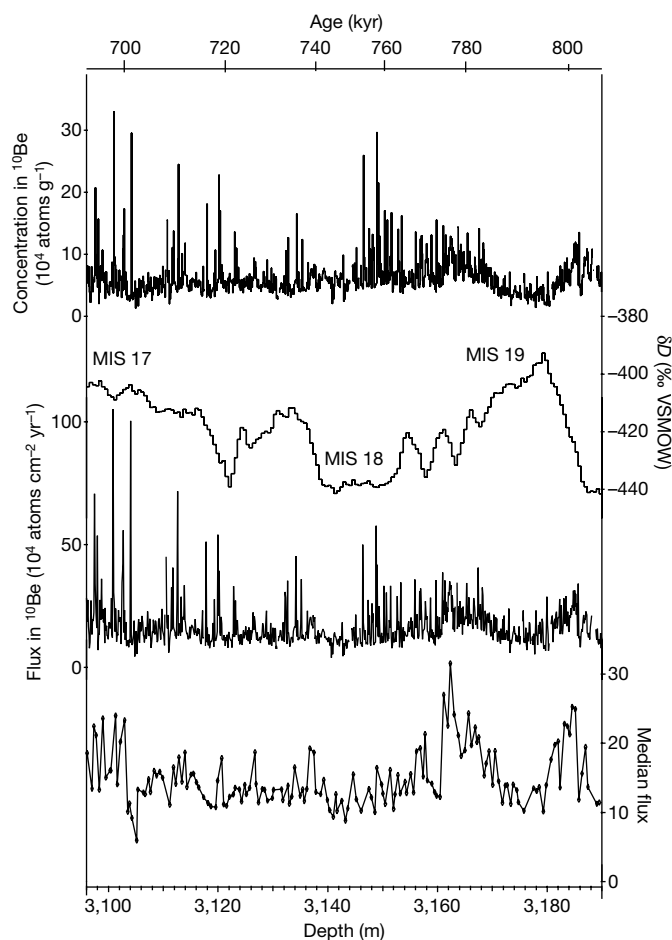


Figure 1 | ^{10}Be concentration, deuterium, ^{10}Be flux and median ^{10}Be flux (see text) in the EPICA Dome C ice core. The ^{10}Be and deuterium (expressed as δD in ‰ with respect to the Vienna standard mean ocean water, VSMOW) are given as a function of depth (bottom scale) and EDC2 age (top scale, see Methods section). Note that because of ice thinning and variable accumulation rates, the age scale is not linear with depth. Intervals believed to correlate with marine isotopic stages MIS 17–19 are shown on the δD curve.

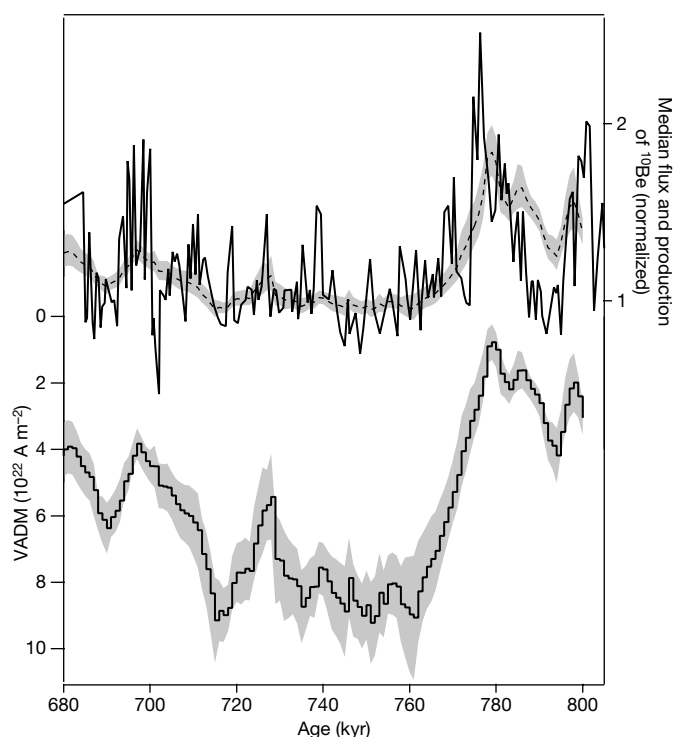


Figure 2 | ^{10}Be flux in EPICA Dome C ice core, geomagnetic dipole intensity from stacked marine sediments¹⁹ and predicted global ^{10}Be production rate. The median ^{10}Be flux (solid line) is from Fig. 1, normalized to the interval 740–767 kyr, a period over which predicted global ^{10}Be production (dashed line), calculated using the model described in ref. 24, is close to unity (that is, the present rate). ^{10}Be from the ice core is plotted on the extended EDC2 timescale. The inferred virtual axial dipole moment (VADM) (solid histogram) with standard error (shaded)¹⁹ and predicted global ^{10}Be production rate (dashed line) with standard error (shaded) are plotted on an astrochronological timescale¹⁹.

no information on how this fraction may have varied in the past. To have an idea of the maximum expected amplitude, however, we can calculate the global variation as a function of the dipole intensity as described in ref. 24, using the estimated dipole intensity shown in Fig. 2. We have carried out this calculation and show it overlaid on the normalized ^{10}Be median flux of Fig. 2. Unexpectedly, the observed increase is slightly larger than the maximum calculated. This suggests that either the calculated sensitivity of the production rate as a function of dipole intensity described in ref. 24 is underestimated, and/or that the actual dipole intensity during the reversal was even smaller than given by the compilation shown in Fig. 2 (the enhancement of the global production rate for a zero dipole intensity in the model of ref. 24 is ~ 2.1). Our observed enhancement for the M–B reversal is also similar to that found in Antarctica during the Laschamp excursion ~ 40 kyr ago^{13,14}. The relatively good agreement between the observed and predicted global enhancements for those two events appears to contradict the idea that the ^{10}Be in Antarctic ice cores reflects primarily local production, at least for those climatic conditions.

In conclusion, we believe the most probable interpretation of the enhanced ^{10}Be flux between 3,160 and 3,170 m in the EPICA Dome C ice core is that it results from the low dipole intensity during the M–B geomagnetic reversal. Although the complication introduced by the observed ^{10}Be spikes prevents us from exploiting the highest-time-resolution palaeointensity information potentially available in this record, if the observed peak is indeed the record of the M–B transition, it suggests that the field remained low and relatively constant for a period of about 12 kyr. A better understanding of the origin of the spikes might allow us to obtain even more detailed information of the variations during a reversal.

If the good correlation between ice core ^{10}Be and relative palaeointensity from marine sediments is found over the whole Brunhes epoch, it would provide independently dated tie points to help constrain ice flow models over this period, as well as an excellent relative common chronology for comparing climatic records from these two types of reservoirs.

METHODS

The ice available for this study was in the form of a continuous series of 55-cm strips (called bag samples), each weighing ~ 300 g. In the depth interval considered here, such samples can represent almost 1,000 yr. To maximize the amount of information from these samples, we have thus decided to divide each one into five pieces, each 11 cm long. The amount of ice per sample (~ 50 g) is almost an order of magnitude smaller than we have used for previous studies^{13,17}. We have therefore modified our previous procedure.

The samples were melted in 250-ml centrifuge cones, in the presence of 0.25 mg of ^9Be carrier. For most of the samples, the $\text{Be}(\text{OH})_2$ was then precipitated directly with NH_4OH . For a few of the deepest samples, a good precipitate did not form, probably because of interference from fluid used to drill the core (our ice strips were taken from the outside of the core). In these cases, we reverted to our previous procedure^{13,17}, in which Be was concentrated on a cation exchange column, eluted with HCl and then precipitated. For all cases, the precipitate was then washed with water (pH = 7), dissolved in a few drops of nitric acid, and transferred to a quartz crucible. The crucible was dried on a hotplate, then heated to 900°C over an electric furnace to transform the precipitate to BeO. The BeO was mixed with Nb powder (325 mesh) in the ratio 3:1 and pressed into a 1-mm-diameter, 1-mm-deep hole in a Mo cathode. This cathode was heated in argon to $1,200^\circ\text{C}$ for 1 min, then $1,800^\circ\text{C}$ for an additional minute, in a resistively heated carbon furnace. This procedure was found to give better and more stable ^9Be currents in the ion source, as well as reduce ^{10}B interference.

The $^{10}\text{Be}/^9\text{Be}$ ratios were measured at the Tandem-based AMS facility in Gif-sur-Yvette relative to the NIST standard SRM 4325, using the certified ratio of 2.68×10^{-11} . For almost all samples, at least 200 ^{10}Be events were recorded, leading to a $1-\sigma$ statistical uncertainty of $<7\%$. Combined with a conservatively estimated 5% machine uncertainty, this leads to an overall uncertainty of $\sim 8.5\%$.

The extended EDC2 timescale was constructed as described in ref. 1, using the deuterium values of Fig. 1 to estimate ice accumulation rates for the ice beyond 740 kyr.

Received 10 February; accepted 18 September 2006.

- EPICA community members. Eight glacial cycles from an Antarctic ice core. *Nature* **429**, 623–628 (2004).
- Jouzel, J. *et al.* Orbital and millennial Antarctic climate variability over the last 800,000 years. *Geophys. Res. Abstr.* **8**, 02821 (2006).
- Coe, R. S., Pringle, M. S. & Zhou, X. Matuyama-Brunhes reversal and Kamikatsura event on Maui: paleomagnetic directions, $^{40}\text{Ar}/^{39}\text{Ar}$ ages and implications. *Earth Planet. Sci. Lett.* **222**, 667–684 (2004).
- Tauxe, L., Herbert, T., Shackleton, N. & Kok, Y. Astronomical calibration of the Matuyama-Brunhes boundary: consequences for magnetic remanence acquisition in marine carbonates and the Asian loess sequences. *Earth Planet. Sci. Lett.* **140**, 133–146 (1996).
- Quidelleur, X., Carlu, J., Solar, V., Valet, J. P. & Gillot, P.-Y. The age and duration of the Matuyama-Brunhes transition from new K-Ar data from La Palma (Canary Islands) and revisited $^{40}\text{Ar}/^{39}\text{Ar}$ ages. *Earth Planet. Sci. Lett.* **208**, 149–163 (2003).
- Hartl, P. & Tauxe, L. A precursor to the Matuyama/Brunhes transition-field instability as recorded in pelagic sediments. *Earth Planet. Sci. Lett.* **138**, 121–135 (1996).
- Brown, L. L., Singer, B. S., Pickens, J. & Jicha, B. R. Paleomagnetic directions and $^{40}\text{Ar}/^{39}\text{Ar}$ ages from the Tatara-San Pedro volcanic complex, Chilean Andes: lava record of a Matuyama-Brunhes precursor? *J. Geophys. Res.* **109**, doi:10.1029/2004JB003007 (2004).
- Singer, B. S. *et al.* Structural and temporal requirements for geomagnetic field reversal deduced from lava flows. *Nature* **434**, 633–636 (2005).
- Merrill, R. T. & McFadden, P. L. Geomagnetic polarity transitions. *Rev. Geophys.* **37**, 201–226 (1999).
- Lal, D. & Peters, B. in *Handbuch der Physik* (ed. Sitte, K.) 551–612 (Springer, New York, 1967).
- Raisbeck, G. M., Yiou, F., Bourles, D. & Kent, D. V. Evidence for an increase in cosmogenic ^{10}Be during a geomagnetic reversal. *Nature* **315**, 315–317 (1985).
- Steig, E. J., Polissar, P. J., Stuiver, M., Grootes, P. M. & Finkel, R. C. Large amplitude solar modulation cycles of ^{10}Be in Antarctica: implications for atmospheric mixing processes and interpretation of the ice core record. *Geophys. Res. Lett.* **23**, 523–526 (1996).
- Raisbeck, G. M. *et al.* Evidence for two intervals of enhanced ^{10}Be deposition in Antarctic ice during the last glacial period. *Nature* **326**, 273–277 (1987).
- Raisbeck, G. M. *et al.* in *Absolute and Radiocarbon Chronologies* (eds Bard, E. & Broecker, W. S.) 127–139 (Springer, Berlin, 1992).
- Yiou, F. *et al.* Beryllium 10 in the Greenland Ice Core Project ice core at Summit Greenland. *J. Geophys. Res.* **102**, 26783–26794 (1997).
- Field, C. V., Schmidt, G. A., Koch, D. & Salyk, C. J. Modeling production and climate-related impacts on ^{10}Be concentration in ice cores. *J. Geophys. Res.* **111**, doi:10.1029/2005JD006410 (2006).
- Yiou, F., Raisbeck, G. M., Bourles, D., Loriaux, C. & Barkov, N. I. ^{10}Be in ice at Vostok Antarctica during the last climatic cycle. *Nature* **316**, 616–617 (1985).
- Durand, G. Microstructure, Recrystallisation et Déformation des Glaces Polaires de la Carotte EPICA, Dôme Concordia, Antarctique. PhD thesis, Université Joseph Fourier, Grenoble (2004).
- Guyodo, Y. & Valet, J. P. Global changes in intensity of the Earth's magnetic field during the past 800 kyr. *Nature* **399**, 249–252 (1999).
- Bassiot, F. C. *et al.* The astronomical theory of climate and the age of the Brunhes-Matuyama magnetic reversal. *Earth Planet. Sci. Lett.* **126**, 91–108 (1994).
- Channell, J. E. T. & Kleiven, H. F. Geomagnetic paleointensities and astrochronological ages for the Matuyama-Brunhes boundary and the boundaries of the Jaramillo Subchron: paleomagnetic records and oxygen isotope records from ODP site 983. *Phil. Trans. R. Soc. Lond. B* **358**, 1027–1047 (2000).
- Hogg, C. S., Palike, H., Wei, K. Y., Liang, W. T. & Torii, M. Astronomically calibrated ages for the geomagnetic reversals with the Matuyama chron. *Earth Planets Space* **54**, 679–690 (2002).
- Langereis, C. G., de Lange, G. J., Paterne, M. & van Santvoort, P. M. J. Magnetostratigraphy and astronomical calibration of the last 1.1 Myr from an eastern Mediterranean piston core and dating of short events in the Brunhes. *Geophys. J. Int.* **129**, 75–94 (1997).
- Wagner, G. *et al.* Reconstruction of the geomagnetic field between 20 and 60 kyr BP from cosmogenic radionuclides in the GRIP ice core. *Nucl. Instrum. Meth. B* **172**, 597–604 (2000).

Acknowledgements We thank S. Jacob, G. Aumaitre, J. Lestringuez, S. Falourd and V. Masson-Delmotte for help with experimental measurements, F. Parrenin for the age and accumulation estimates, and K. Krogh-Andersen for suggesting the use of medians to eliminate the effect of the ^{10}Be spikes. Tandem operation is supported by the IN2P3 and INSU divisions of the CNRS. This work is a contribution to the European Project for Ice Coring in Antarctica (EPICA), a joint European Science Foundation/European Commission (EC) scientific programme, funded by the EC and by national contributions from Belgium, Denmark, France, Germany, Italy, The Netherlands, Norway, Sweden, Switzerland and the UK. The main logistic support was provided by IPEV and PNRA.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.R. (raisbeck@csnsm.in2p3.fr).

Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida

Sarah J. Boulrat¹, Thorhildur Juliusdottir², Christopher J. Lowe³, Robert Freeman⁴, Jochanan Aronowicz³, Mark Kirschner⁵, Eric S. Lander^{4,6}, Michael Thorndyke⁷, Hiroaki Nakano⁷, Andrea B. Kohn⁸, Andreas Heyland⁸, Leonid L. Moroz⁸, Richard R. Copley² & Maximilian J. Telford¹

Deuterostomes comprise vertebrates, the related invertebrate chordates (tunicates and cephalochordates) and three other invertebrate taxa: hemichordates, echinoderms and *Xenoturbella*¹. The relationships between invertebrate and vertebrate deuterostomes are clearly important for understanding our own distant origins. Recent phylogenetic studies of chordate classes and a sea urchin have indicated that urochordates might be the closest invertebrate sister group of vertebrates, rather than cephalochordates, as traditionally believed^{2–5}. More remarkable is the suggestion that cephalochordates are closer to echinoderms than to vertebrates and urochordates, meaning that chordates are paraphyletic². To study the relationships among all deuterostome groups, we have assembled an alignment of more than 35,000 homologous amino acids, including new data from a hemichordate, starfish and *Xenoturbella*. We have also sequenced the mitochondrial genome of *Xenoturbella*. We support the clades Olfactores (urochordates and vertebrates) and Ambulacraria (hemichordates and echinoderms⁶). Analyses using our new data, however, do not support a cephalochordate and echinoderm grouping and we conclude that chordates are monophyletic. Finally, nuclear and mitochondrial data place *Xenoturbella* as the sister group of the two ambulacrarian phyla¹. As such, *Xenoturbella* is shown to be an independent phylum, Xenoturbellida, bringing the number of living deuterostome phyla to four.

Although molecular phylogenetic studies have prompted changes in the membership of the deuterostome superphylum (two groups—chaetognaths⁷ and lophophorates⁸—have been reclassified as protozoans and one species, *Xenoturbella bocki*, added¹) there remains an unchanging core of three phyla: echinoderms, hemichordates and chordates. Three main groups are recognized within the chordates themselves—urochordates, including the ascidians and lancelets; cephalochordates or lancelets; and vertebrates, including fish and tetrapods.

Molecular studies have also changed our view of the relationships between deuterostome taxa. Hemichordates, long thought of as the sister group of the chordates, are now considered to be closest to echinoderms in a supraphyletic clade named the Ambulacraria⁶. Most recently, analyses of hundreds of genes from expressed sequence tag (EST) and whole-genome sequencing projects have indicated that, contrary to previous studies, urochordates are the invertebrate group that is most closely related to vertebrates². Independent molecular^{3,4} and morphological⁹ support for this unexpected grouping also exists. Although Ambulacraria and Olfactores

(the urochordate plus vertebrate clade¹⁰) seem to be credible components of the deuterostomes, two further aspects of the phylogeny of the group remain contentious.

The first stems from a remarkable result in a phylogenomic analysis in which cephalochordates were associated with echinoderms rather than with other chordates². This result, if correct, has profound implications for our understanding of early deuterostome evolution and the origins of the chordate body plan, because it suggests that a number of conserved chordate features were present in the common ancestor of all deuterostomes and were secondarily lost in echinoderms and hemichordates. Although this is clearly a controversial finding, it deserves serious consideration. The initially surprising relocation of the hemichordates from chordate sister group to the Ambulacraria has already shown that certain characteristics that were thought to be specific to the chordate lineage are likely to derive from the deuterostome common ancestor; the homology of gill slits and endostyle in hemichordates and chordates is now supported by studies both of morphology and of shared gene-expression patterns^{9,11,12}.

The second open question concerns the position of the worm *Xenoturbella*. Consideration of its small subunit ribosomal RNA gene (SSU) indicated that *Xenoturbella* might be a deuterostome, related to the Ambulacraria but not included in either the echinoderms or the hemichordates¹. This result is supported by the lack in *Xenoturbella* of a complex molecular character that unites the two ambulacrarian phyla—a change in their mitochondrial genetic code whereby the codon ATA codes for the amino acid isoleucine rather than for methionine as in most other animals, including *Xenoturbella*¹. At the same time, there are several reasons to question the position of *Xenoturbella*. There is low statistical support for its exclusion from the hemichordates in the SSU study¹, the change in genetic code is known to be homoplastic because it has occurred independently in the rhabditophoran platyhelminths¹³; separate analyses of mitochondrial *Cytochrome oxidase I* sequences placed *Xenoturbella* with the hemichordates¹, and there are detailed similarities between the structure of the epidermis of hemichordates and *Xenoturbella*¹⁴ that have led to suggestions that *Xenoturbella* is a neotenous hemichordate¹⁵.

To test the robustness of previous work and to extend the phylogenomic approach to all of the main deuterostome groups, we have gathered data from the two groups previously missing: a hemichordate, *Saccoglossus kowalevski* (206,715 ESTs including those from Lowe *et al.*¹⁶) and *Xenoturbella bocki* (1,372 ESTs). We added to

¹Department of Biology, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK. ²Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK. ³Department of Organismal Biology and Anatomy, 1027 E. 57th Street, University of Chicago, Chicago, Illinois 60637, USA. ⁴Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA. ⁵University of California, Berkeley, Department of Molecular and Cell Biology, 142 Life Sciences Addition #3200, Berkeley, California 94720-3200, USA. ⁶Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ⁷Royal Swedish Academy of Sciences, Kristineberg Marine Research Station, Fiskebäckskil, S 450 34, Sweden. ⁸The Whitney Laboratory for Marine Bioscience, University of Florida, 9505 Ocean Shore Boulevard, St Augustine, Florida 32080, USA.

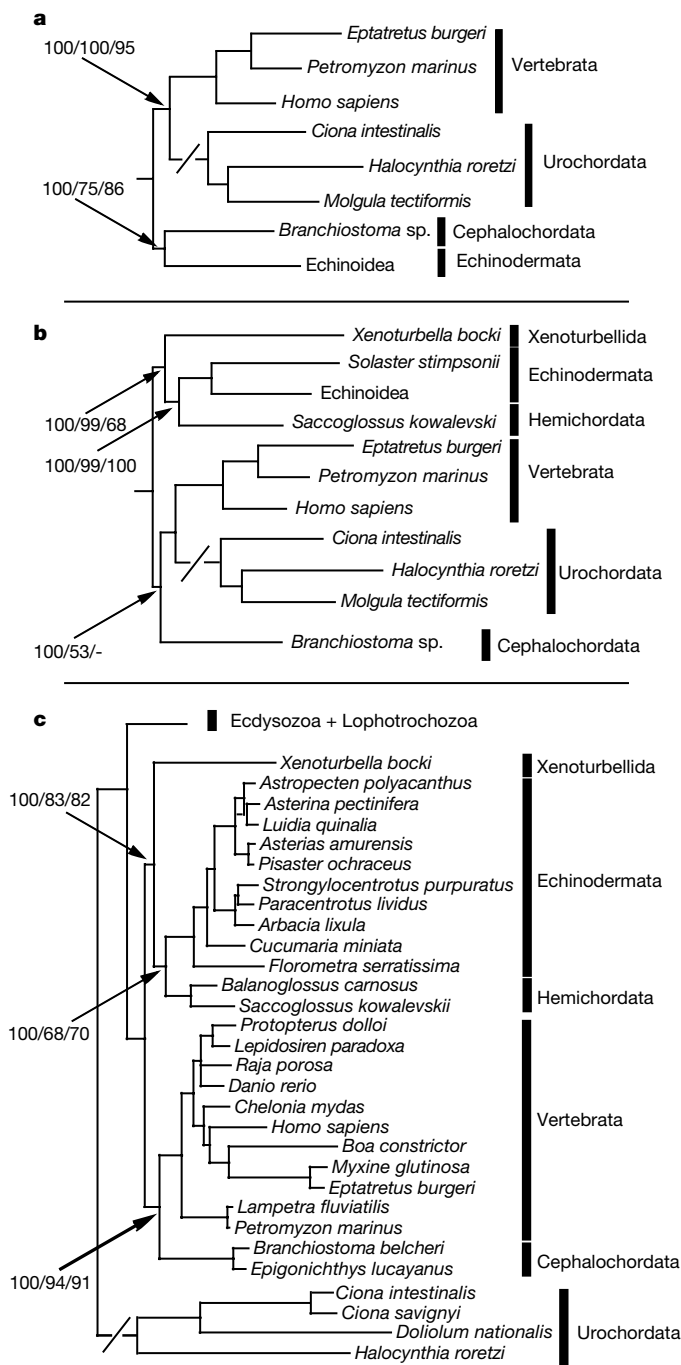


Figure 1 | Phylogenetic analyses of 170 nuclear proteins and 13 mitochondrial proteins support a monophyletic chordate clade and an independent deuterostome phylum of Xenoturbellida. **a**, Bayesian analysis of nuclear data before the addition of the new asteroid, hemichordate and xenoturbellid data. The cephalochordate groups with the echinoderm, implying that the chordates are paraphyletic. **b**, Bayesian analysis of nuclear data after the addition of asteroid, hemichordate and xenoturbellid data. The new sequences join the branch to the echinoderm, and the cephalochordates now join the chordate branch. This indicates that the previous result is due to systematic error. *Xenoturbella* is the sister group of the Ambulacraria (echinoderms plus hemichordates). **c**, Bayesian analysis of mitochondrial data with the amino acids M, I, N and K recoded as missing data places the cephalochordates with vertebrates; *Xenoturbella* is the sister group of Ambulacraria. Support for critical nodes is shown in the order BPP (%), TF (%) and MLBP (%). Long branches leading to the urochordates have been shortened as indicated by a break (/) in the branch (see Supplementary Information for unaltered tree and scale).

the sampling of the echinoderm clade with data from the sun star *Solaster stimpsonii* (5,614 ESTs). We searched these ESTs for orthologues of the 146 genes used by Philippe *et al.*¹⁷. We then supplemented these with 35 genes having a representative orthologue in our *Xenoturbella* ESTs, as this taxon had the fewest matches to the data set from ref. 17. We culled 11 alignments where orthology was uncertain, leaving 170 concatenated, aligned orthologous genes (35,588 reliably aligned amino acids). *Saccoglossus* had data from 153 of 170 genes and 29,777 amino acids, *Xenoturbella* had 63 genes and 8,370 amino acids and *Solaster* had 67 genes and 11,033 amino acids.

We compared results from our extended data set of 170 genes to those of the 146-gene data set of Delsuc *et al.*² by performing phylogenetic analyses without the new hemichordate, *Xenoturbella* and sun star sequences. Phylogenetic analyses using maximum-likelihood and bayesian approaches recapitulated previous results (Fig. 1a). All nodes received maximum support from bayesian analysis (bayesian posterior probability (BPP) = 100%). Maximum-likelihood non-parametric bootstrap support (MLBP) was lower for some nodes (see Supplementary Information). The overall topology of this tree conforms to the 'new animal phylogeny'¹⁸ with monophyletic protostomes, deuterostomes, lophotrochozoans (represented here by molluscs and an annelid), and ecdysozoans (represented here by arthropods and a nematode; see Supplementary Information). Within the deuterostomes, we find support for monophyletic groups of vertebrates, cyclostomes (hagfish and lamprey) and Olfactores (vertebrates plus urochordates). Importantly, we recovered the relationship linking cephalochordates not to other chordates but to echinoderms: BPP = 100%, Treefinder¹⁹ LRSH support (TF) = 75% and MLBP = 81%. This relationship was unexpected and we wanted to test the possibility that it was caused by systematic error. Two powerful ways of addressing systematic errors are to include additional taxa, thereby enabling better detection of misleading multiple substitutions, and to remove problematic taxa (for example, long branches); we have used both approaches²⁰.

By adding sequences from three taxa that are likely to be related to the echinoids (asteroid, hemichordate and *Xenoturbella*) we could test the possibility that the cephalochordate/echinoderm relationship might result from a systematic error such as long-branch attraction²¹. The effect of including these three taxa in our analysis is clear; support for a cephalochordate plus echinoderm clade disappears and support for the canonical monophyletic chordates increases (Fig. 1b). The monophyletic chordate clade in our new analysis has BPP = 100% and TF = 53%, but no MLBP support. This change in topology on the addition of taxa that divide the echinoid branch is consistent with the idea that the attraction between the cephalochordate and the echinoid results from a systematic error; the addition of non-problematic taxa tends to improve the reliability of a phylogeny²¹. We compared the support for the optimal topology to that for the alternative topology, which groups the cephalochordates with Ambulacraria and *Xenoturbella* using the Shimodaira–Hasegawa (SH) tests and Bayes factors. The SH test did not reject the alternative topology as significantly worse than the optimal topology (SH $P = 0.476$), but the alternative topology was rejected as significantly worse using the Bayes factors test ($2\log_e(B10) = 24$), although this test might not be conservative. The inclusion of data from the other major taxon of cephalochordates—the Epigonichthyidae—will divide the cephalochordate branch and, we predict, will further increase support for the monophyletic chordates.

A parallel possibility is that rapidly evolving urochordates, along with their sister group the vertebrates, have been attracted towards the base of the tree by long-branch effects²². Two lines of evidence support this possibility. First, in the analyses described above, we used diploblast metazoans (ctenophore and cnidarian) as close outgroups; the inclusion of more distant outgroups (yeast and choanoflagellates), as in ref. 2, results in a more basal position for Olfactores with the corollary that cephalochordates group with echinoderms. This result is observed even when the asteroid, hemichordate

Table 1 | Proportions of amino acids affected by mitochondrial genetic code change in echinoderms and other metazoans

	M	I	N	K
Echinoderm	2.32 (0.09)	10.28 (1.03)	7.05 (0.64)	1.43 (0.02)
Metazoan	5.51 (2.31) **	8.1 (2.6) **	3.52 (0.66) n.s.	4.14 (0.20) **
Deuterostome	5.31 (1.34) **	7.79 (1.25) n.s.	3.58 (0.24) **	3.96 (0.28) **

Mean and variance (in parentheses) of percentage proportions of the amino acids methionine (M), isoleucine (I), asparagine (N) and lysine (K) in the mitochondrial genes of 12 echinoderms, 70 metazoans and 23 non-echinoderm deuterostomes. *F*-test two-sample for variance test was used to compare metazoan and deuterostome values to those of the echinoderms. ** indicates $P < 0.05$. n.s. indicates non-significant difference.

and *Xenoturbella* data are included. Second, exclusion of the long-branched urochordates always results in cephalochordates grouping with vertebrates with or without our additional taxa and even in the presence of distant outgroups (see Supplementary Information).

We also considered the phylogeny of the other deuterostome groups. The hemichordate is the sister group of the echinoderms (asteroid and echinoid), which gives weight to a monophyletic Ambulacraria. Of greater interest is our finding that *Xenoturbella* is related to but outside the monophyletic Ambulacraria. The clade (*Xenoturbella* plus Ambulacraria) has BPP = 100%, TF = 99% and MLBP = 68% whereas ambulacrarian monophyly excluding *Xenoturbella* has BPP = 100%, TF = 99.3% and MLBP = 100%. We compared the support for the optimal topology to that for alternative possibilities in which *Xenoturbella* is grouped with hemichordates or echinoderms using the SH tests and Bayes factors. Both tests rejected as significantly less well supported the grouping of *Xenoturbella* with hemichordates (SH $P = 0.012$, Bayes factors $2\log_e(B10) = 96$) or *Xenoturbella* with echinoderms (SH $P = 0.037$, Bayes factors $2\log_e(B10) = 118$).

To test our findings further, we sequenced the entire mitochondrial genome (15,234 nucleotides) of *Xenoturbella bocki*. We have used the amino-acid sequences of all 13 protein-coding loci to examine the position of *Xenoturbella* within the deuterostomes. Our initial results placed *Xenoturbella* at the base of the deuterostomes rather than as the sister group of Ambulacraria. We were aware, however, that the mitochondrial genetic code changes in the lineage leading to Ambulacraria and an additional change in the echinoderms¹³ might have led to compositional bias that could artefactually distance them from *Xenoturbella*. Indeed, the proportions of amino acids that were affected by the changes of genetic code (methionine and isoleucine, asparagine and lysine (M, I, N, K)) in the echinoderms differ from the average seen in deuterostomes and metazoans (Table 1). We tested the effects of this imbalance on the position of *Xenoturbella* in two ways. First we excluded the most biased taxa—the echinoderms—leaving the hemichordates as representatives of Ambulacraria. The result of excluding the echinoderms is to move *Xenoturbella* next to the hemichordates, as predicted if its basal position were due to compositional bias in the echinoderms (see Supplementary Information). Second, we avoided the influence of the problematic amino acids (M, I, N, K) by recoding each instance of them as 'missing data', which affected 33.9% of our reliably aligned amino acids. The effect of this was again to move *Xenoturbella* to become the sister group of Ambulacraria. *Xenoturbella* plus Ambulacraria monophyly has BPP = 100%, TF = 83% and MLBP = 82% whereas ambulacrarian monophyly excluding *Xenoturbella* has BPP = 100%, TF = 68% and MLBP = 70% (Fig. 1c).

We note that according to our mitochondrial data set, the cephalochordates and vertebrates form a highly supported monophyletic group (BPP = 100%, TF = 96.5% and MLBP = 91%), which further diminishes the credibility of the cephalochordate–echinoderm link. In common with previous analyses²³, our tree reconstructions place the long-branched urochordates as a sister group to all other Bilateria.

What conclusions can we draw from this knowledge of the relationships of all the main deuterostome clades? Our analyses indicate that the relationship between cephalochordates and echinoderms

probably results from systematic error. This makes excellent sense from a morphological standpoint because chordates share several characteristics (somites, dorsal nerve chord, notochord and hypophysis) that are absent from Ambulacraria, *Xenoturbella* and non-deuterostome phyla²⁴. Although hemichordates have in the past been considered to possess homologues of the chordate notochord and dorsal nerve cord, these similarities do not stand up to close scrutiny on the basis of ultrastructure or shared expression of orthologous patterning genes^{9,16}.

We have also shown compelling evidence that *Xenoturbella* should be positioned within the deuterostomes as a sister group to Ambulacraria. This result has important consequences for our understanding of deuterostome evolution. First, it means that *Xenoturbella* is not a neotenous member of the hemichordates¹⁵ but an independent lineage—*Xenoturbellida*²⁵—which constitutes a newly recognized deuterostome phylum. We propose the name Xenambulacraria for the highly supported monophyletic grouping of *Xenoturbellida* plus Ambulacraria.

The second implication of this result is that characteristics shared by the hemichordates and *Xenoturbella* can be parsimoniously reconstructed as having been present in the last common ancestor of Xenambulacraria. Characteristics of this ancestor must have included those characteristics shared by all deuterostomes (deuterostomy, radial cleavage, enterocoely); those shared by hemichordates and chordates (gill slits and endostyle); and the basiepithelial nervous system^{15,26,27} and unusual ciliary structure shared by the hemichordates and *Xenoturbella*^{28,29}. Evidence for most of these characteristics remains to be found in the little-studied *Xenoturbella*.

Although our results, which support a monophyletic chordate clade, conform to a reasonable view of unique chordate characters, we are, nevertheless, returned to a state of relative ignorance concerning the characteristics and evolution of the earliest deuterostomes. Perhaps the biggest question is whether the ancestral deuterostome nervous system was centralized, as in chordates, or diffuse, as in Xenambulacraria. The central nervous system that is prevalent in the protostomes—the closest sister group of the deuterostomes—suggests the former, but a diffuse nervous system is found in the cnidarians and acelomorph flatworms, which are widely thought to be the sister group of the protostomes plus deuterostomes³⁰. Further progress in understanding the origins and evolution of deuterostome and chordate characteristics is likely to come from the integration of palaeontological, comparative embryological and genomic approaches.

METHODS

Data set assembly. Representative sequences from the data set of Philippe *et al.*¹⁷ were searched against the est_others subset of GenBank, supplemented with EST sequences from *Saccoglossus* (206,715 sequences), *Solaster* (5,614 sequences) and *Xenoturbella* (1,372 sequences) using tblastn. Hits from our EST data sets plus those from the lamprey *Petromyzon marinus*, hagfish *Eptatretus burgeri* and the tunicates *Molgula tectiformis* and *Oikopleura dioica* were then assembled and added to these alignments. Additional data sets were assembled using a starting set of *Xenoturbella* proteins that had an orthologue in *Homo*, *Drosophila* and *Caenorhabditis* (see Supplementary Information). The *Xenoturbella* mitochondrial genome was sequenced as described in the Supplementary Information. Additional mitochondrial genomes came from the OGRE website (drake.physics.mcmaster.ca/ogre/index.shtml).

Phylogenetic analyses. The concatenated amino-acid alignments were analysed using bayesian and maximum-likelihood approaches and maximum-likelihood

non-parametric bootstrapping. Likelihood-based tests of alternative topologies used SH tests and Bayes factors (see Supplementary Information).

Received 8 June; accepted 11 September 2006.

Published online 18 October 2006.

- Bourlat, S. J., Nielsen, C., Lockyer, A. E., Littlewood, D. T. J. & Telford, M. J. *Xenoturbella* is a deuterostome that eats molluscs. *Nature* **424**, 925–928 (2003).
- Delsuc, F., Brinkmann, H., Chourrout, D. & Philippe, H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439**, 965–968 (2006).
- Oda, H. *et al.* A novel amphioxus cadherin that localises to epithelial adherens junctions has an unusual domain organisation with implications for chordate phylogeny. *Evol. Dev.* **4**, 426–434 (2002).
- Wada, H., Okuyama, M., Satoh, N. & Zhang, S. Molecular evolution of fibrillar collagen in chordates, with implications for the evolution of vertebrate skeletons and chordate phylogeny. *Evol. Dev.* **8**, 370–377 (2006).
- Jefferies, W. R., Strickler, A. G. & Yamamoto, Y. Migratory neural crest-like cells form body pigmentation in a urochordate embryo. *Nature* **431**, 696–699 (2004).
- Turbeville, J. M., Schulz, J. R. & Raff, R. A. Deuterostome phylogeny and the sister group of the chordates: evidence from molecules and morphology. *Mol. Biol. Evol.* **11**, 648–655 (1994).
- Telford, M. J. & Holland, P. W. H. The phylogenetic affinities of the chaetognaths: a molecular analysis. *Mol. Biol. Evol.* **10**, 660–676 (1993).
- Halanych, K. M. *et al.* Evidence from 18S ribosomal DNA that the lophophorates are protostome animals. *Science* **267**, 1641–1643 (1995).
- Ruppert, E. E. Key characters uniting hemichordates and chordates: homologies or homoplasies? *Can. J. Zool.* **83**, 8–23 (2005).
- Jefferies, R. P. S. in *Biological Asymmetry and Handedness* (eds Bock, G. R. & Marsh, J.) 94–127 (John Wiley, Chichester, 1991).
- Ogasawara, M., Wada, H., Peers, H. & Satoh, N. Developmental expression of Pax 1/9 genes in urochordate and hemichordate gills: insight into function and evolution of the pharyngeal epithelium. *Development* **126**, 2539–2550 (1999).
- Rychel, A. L., Smith, S. E., Shimamoto, H. T. & Swalla, B. J. Evolution and development of the chordates: collagen and pharyngeal cartilage. *Mol. Biol. Evol.* **23**, 541–549 (2006).
- Telford, M. J., Herniou, E. A., Russell, R. B. & Littlewood, D. T. J. Changes in mitochondrial genetic codes as phylogenetic characters: two examples from the flatworms. *Proc. Natl Acad. Sci. USA* **97**, 11359–11364 (2000).
- Stach, T. *et al.* Nerve cells of *Xenoturbella* (phylum uncertain) and *Harrimania kupfferi* (Enteropneusta) are positively immunoreactive to antibodies raised against echinoderm neuropeptides. *J. Mar. Biol. Assoc. UK* **85**, 1519–1524 (2005).
- Reisinger, E. Was ist *Xenoturbella*? *Z. Wissenschaftliche Zool.* **164**, 188–198 (1960).
- Lowe, C. J. *et al.* Anteroposterior patterning in hemichordates and the origins of the chordate nervous system. *Cell* **113**, 853–865 (2003).
- Philippe, H., Lartillot, N. & Brinkmann, H. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol. Biol. Evol.* **22**, 1246–1253 (2005).
- Aguinaldo, A. M. A. & Lake, J. A. Evolution of the multicellular animals. *Am. Zool.* **38**, 878–887 (1998).
- Jobb, G. TREEFINDER version of May 2006. (<http://www.treefinder.de>) (2006).
- Phillipe, H. & Telford, M. J. Large-scale sequencing and the new animal phylogeny. *Trends Ecol. Evol.* (in the press).
- Wiens, J. J. Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? *Syst. Biol.* **54**, 731–742 (2005).
- Felsenstein, J. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27**, 401–410 (1978).
- Yokobori, S., Watanabe, Y. & Oshima, T. Mitochondrial genome of *Ciona savignyi* (Urochordata, Ascidiacea, Enterogona): comparison of gene arrangement and tRNA genes with *Halocynthia roretzi* mitochondrial genome. *J. Mol. Evol.* **57**, 574–587 (2003).
- Weichert, C. K. *Anatomy of the Chordates* (McGraw-Hill, New York, 1965).
- Westblad, E. *Xenoturbella bocki* n.g., n.sp., a peculiar, primitive turbellarian type. *Ark. Zool.* **1**, 3–29 (1949).
- Pedersen, K. J. & Pedersen, L. R. Ultrastructural observations on the epidermis of *Xenoturbella bocki* Westblad, 1949—with a discussion of epidermal cytoplasmic filament systems of invertebrates. *Acta Zool.* **69**, 231–246 (1988).
- Raikova, O. I., Reuter, M., Jondelius, U. & Gustafsson, M. K. S. An immunocytochemical and ultrastructural study of the nervous and muscular systems of *Xenoturbella westbladi* (Bilateria inc. sed.). *Zoomorphology* **120**, 107–118 (2000).
- Franzen, A. & Afzelius, B. A. The ciliated epidermis of *Xenoturbella bocki* (Platyhelminthes, Xenoturbellida) with some phylogenetic considerations. *Zool. Scr.* **16**, 9–17 (1987).
- Pardos, F. Fine structure and function of pharynx cilia in *Glossobalanus minutus* Kowalewsky (Entropneusta). *Acta Zool.* **69**, 1–12 (1988).
- Ruiz Trillo, I., Riutort, M., Littlewood, D. T. J., Herniou, E. A. & Baguña, J. Acoel flatworms: earliest extant bilaterian metazoans, not members of Platyhelminthes. *Science* **283**, 1919–1923 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements R.R.C., T.J., M.J.T. and S.J.B. are supported by the EU Marie Curie RTN Zoonet. S.J.B. is also funded by the BBSRC. H.N. was supported by the Human Frontier Science Program Long-Term Fellowship. L.L.M. was supported by NIH and NSF grants as well as the McKnight Brain Research Foundation and the University of Florida opportunity funds.

Author Contributions S.J.B. sequenced the mitochondrial genome. C.J.L., R.F., J.A., M.K. and E.S.L. generated the *Saccoglossus* and *Solaster* EST data. M.T., H.N., A.B.K., A.H. and L.L.M. generated the *Xenoturbella* EST data. R.R.C. and T.J. assembled the EST data. M.J.T. analysed the data and led the write-up.

Author Information The *Xenoturbella* mitochondrial genome sequence has GenBank Accession number DQ832701. *Xenoturbella* ESTs have accession numbers EC906293–EC907475 and novel *Saccoglossus* ESTs used in this project have accession numbers EE111315–EE122968. Novel *Solaster* ESTs used in this project have accession numbers EE122969–EE123339. Alignments used are available for download as Supplementary Information or on request from M.J.T. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.J.T. (m.telford@ucl.ac.uk).

Post-mating sexual selection increases lifetime fitness of polyandrous females in the wild

Diana O. Fisher¹, Michael C. Double¹, Simon P. Blomberg², Michael D. Jennions¹ & Andrew Cockburn¹

Females often mate with several males before producing offspring¹. Field studies of vertebrates suggest, and laboratory experiments on invertebrates confirm, that even when males provide no material benefits, polyandry can enhance offspring survival^{2,3}. This enhancement is widely attributed to genetic benefits that arise whenever paternity is biased towards males that sire more viable offspring^{1,4,5}. Field studies suggest that post-mating sexual selection biases fertilization towards genetically more compatible males^{6,7} and one controlled experiment has shown that, when females mate with close kin, polyandry reduces the relative number of inbred offspring⁸. Another potential genetic benefit of polyandry is that it increases offspring survival because males with more competitive ejaculates sire more viable offspring⁹. Surprisingly, however, there is no unequivocal evidence for this process¹⁰. Here, by experimentally assigning mates to females, we show that polyandry greatly increases offspring survival in the Australian marsupial *Antechinus stuartii*. DNA profiling shows that males that gain high paternity under sperm competition sire offspring that are more viable. This beneficial effect occurs in both the laboratory and the wild. Crucially, there are no confounding non-genetic maternal effects that could arise if polyandry increases female investment in a particular reproductive event¹⁰ because *A. stuartii* is effectively semelparous. Our results therefore show that polyandry improves female lifetime fitness in nature. The threefold increase in offspring survival is not negated by a decline in maternal lifespan and is too large to be offset by an equivalent decline in the reproductive performance of surviving offspring.

Field studies show that the offspring of polyandrous females often have lower juvenile mortality². Experiments that manipulate the number of mates while controlling for mating frequency confirm that this is directly due to polyandry^{3,10}. These experiments are largely restricted to laboratory studies of invertebrates¹⁰. It is unclear whether polyandry has an equivalent effect in the field, because juvenile survival depends heavily on the interaction between investment in growth and environmental conditions¹¹. More importantly, these experiments involve species with a breeding biology that makes it difficult to distinguish between competing explanations for how polyandry improves offspring survival or, ultimately, net fitness.

First, females paired with preferred males often invest more in reproduction¹². Similar differential allocation could occur if females increase investment in reproductive bouts when they have more mates¹⁰. These maternal effects could account for the higher offspring survival of polyandrous females. Four studies that directly tested for maternal effects showed that these effects contributed to polyandry improving offspring performance^{11,13–15}. This makes it problematic to conclude that polyandry is beneficial solely due to genetic benefits arising from greater fertilization by sperm carrying genes that increase juvenile survival. Second, paternity tests on older offspring

could create a spurious correlation between estimated fertilization success and offspring viability if early mortality varies among the progeny of competing males¹⁰. One should therefore assess paternity shortly after fertilization, or use mortality rates of offspring of known parentage to correct, retrospectively, initial estimates of paternity. Third, polyandry confers genetic benefits only if it increases net offspring fitness. There is evidence that males with more competitive ejaculates sire offspring with higher values for specific fitness components, but these components do not include survivorship^{16,17}. More importantly, there is no clear evidence linking net offspring fitness to ejaculate competitiveness because there are plausible life history trade-offs with unmeasured fitness components¹.

Here we have investigated how polyandry affects offspring survival in the brown antechinus *Antechinus stuartii*. The unusual life history of this Australian marsupial enables us to resolve these problems. The annual mating season lasts 10–14 d, during which females mate multiply and then give birth synchronously. Paternal care is absent because males die before offspring are born¹⁸. Crucially, ~93% of females breed once because survival between years is extremely low. Females should therefore invest maximally in every litter. Gestation lasts ~27 d, but embryos remain in a suspended state and implant only 4–5 d before birth¹⁹. Unlike placental mammals, young are born as early stage embryos (mass, ~0.016 g) and fuse obligatorily to a teat for 35–40 d before detaching. Young are weaned after ~90 d (ref. 20). The average female produces 8% more young than can attach to her 8–10 teats and the excess die¹⁹. We can assign paternity and monitor survival from birth.

We conducted two separate experiments. In 2003, we trapped antechinuses from a natural population shortly before the mating period. In the laboratory, we assigned 17 females to a polyandry (three males per female) and 19 to a monandry (one male per female, three times) mating treatment. The proportion of females with young attached to every teat (see Supplementary Information) was higher for polyandrous females ($\chi^2 = 8.0$, $P = 0.005$). There was no early offspring mortality before releasing females at the original study site after ~35 d; however, the proportion of each litter that then survived to weaning was threefold greater for polyandrous females ($\chi^2 = 21.9$, $P < 0.0001$; Fig. 1). There was no subsequent effect of mating treatment on the survival of the remaining offspring to the next breeding season (Jolly–Seber estimates at 4-week intervals: polyandrous, 0.83 ± 0.05 (mean $\pm$ s.e.m.); monandrous, 0.84 ± 0.05).

In 2004, we replicated the experiment but kept lactating females in captivity until just before weaning. We therefore determined whether the previous benefit of polyandry depended on stressful natural conditions, or was replicable in a benign laboratory environment (*ad libitum* food). We also calculated the relationship between the share of paternity of a male and the survival of his offspring. We used independent data (different females) to measure each parameter in a blocked design with 48 wild-caught females and 24 wild-caught

¹School of Botany and Zoology, and ²Centre for Resource and Environmental Studies, Australian National University, Canberra, Australian Capital Territory 0200, Australia.

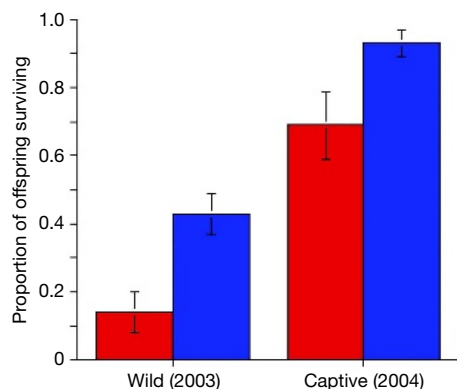


Figure 1 | Effect of mating treatment and rearing environment on the proportion of offspring surviving to weaning. Survival in the wild is estimated from the number of offspring that were captured at weaning. We released 100 offspring from 17 monandrous females and 139 offspring from 19 polyandrous females into the wild. Survival in captivity is calculated from the number of offspring that survived until weaning. There were 121 offspring from 18 monandrous females and 116 offspring from 18 polyandrous females in the captive experiment. Red bars indicate monandry; blue bars indicate polyandry. Data are the mean $\pm$ s.e.m.

males. In a block, each of three males mated once to three different polyandrous females (systematically varying male mating order) and three times with one of three monandrous females. Although the effect of mating treatment did not differ between years (Likelihood ratio test: $\chi^2 = 2.5$, $P = 0.11$), polyandry had no effect on the proportion of females that had young attached to every teat in 2004 ($\chi^2 = 0.54$, $P = 0.46$). The effect of polyandry on the proportion of females that had young attached to every teat is therefore equivocal (both years: polyandry, 0.59 ± 0.08 ; monandry, 0.32 ± 0.07).

Captive females had significantly lower offspring mortality than those released into the wild ($\chi^2 = 146.2$, $P \ll 0.001$); however, offspring survival to weaning was still higher for polyandrous females ($\chi^2 = 4.0$, $P = 0.046$; Fig. 1). In fact, the offspring viability benefit of polyandry was the same for captive and wild females ($\chi^2 = 0.4$, $P = 0.53$), so there was no detectable effect of the rearing environment. The mating treatment effect was due to a sudden increase in offspring mortality for monogamous females after 65–70 d (Fig. 2).

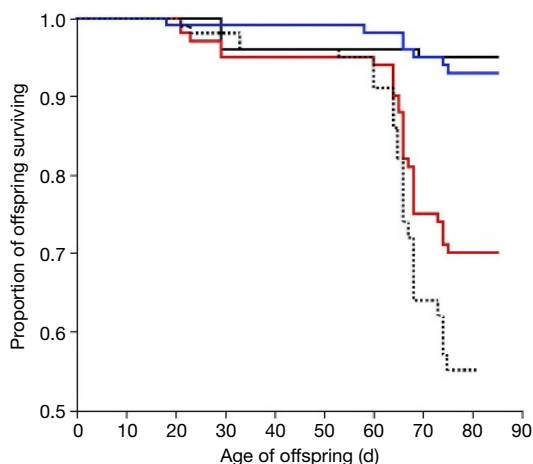


Figure 2 | Effect of mating treatment and male ejaculate competitiveness on survival of offspring in captivity from birth to release (2004). Mean survival curves are presented for monandrous females mated to more competitive males ($n = 41$ offspring from five litters; unbroken black line) or less competitive males ($n = 80$ offspring from 13 litters; dotted black line) and polyandrous females ($n = 116$ offspring from 18 litters; blue line). The pooled curve for all 18 monandrous females that gave birth is also shown (red line).

We counted the total number of offspring that each male sired with three polyandrous females to estimate his ejaculate competitiveness. The survival of offspring produced when females mated monandrously increased strongly with the number of offspring sired by their mate in the polyandrous treatment ($\chi^2 = 36.8$, $P \ll 0.001$). We also divided males into two categories: those with less competitive and more competitive ejaculates (0–7 versus 9–15 offspring; Fig. 3). Females that were assigned a more competitive male did not have young on every teat more often than those assigned a less competitive male ($\chi^2 = 2.1$, $P = 0.15$). However, offspring survival to weaning was far higher for females that mated with a more competitive than a less competitive male (Wald test: $\chi^2 = 4.20$, $P = 0.019$), and no different to that of polyandrous females ($\chi^2 = 0.002$, $P = 0.49$; Fig. 2).

This study shows that genetic benefits from polyandry increase female fitness in nature in the absence of male provisioning effects and that this effect can be replicated in captivity. We can eliminate confounding non-genetic maternal effects because semelparity disfavours deferment of reproductive investment, and offspring actually died while mothers were still freely lactating. The threefold increase in the probability of offspring surviving to weaning is unlikely to be negated by unmeasured fitness costs. We can dismiss an effect of polyandry on lifetime litter production as most females breed once²⁰, and mating treatment did not affect maternal survival between years (Fisher's exact test, $P = 0.81$). There was also no detectable trade-off between early and late offspring mortality as polyandrous and monogamous females had almost identical post-weaning offspring survivorship. Although we did not directly determine offspring lifetime fitness, it is highly improbable that unmeasured life history trade-offs are sufficiently strong to eliminate the threefold increase in offspring survival. A negative effect of polyandry on offspring reproductive success is unlikely. Male success is determined largely by body size^{18,21,22} and female fecundity in this study was unrelated to body size ($P = 0.98$, $n = 86$). As seen in a closely related species²³, the offspring of polyandrous females were actually slightly bigger ($\chi^2 = 6.2$, $P = 0.01$; see Supplementary Information). A possible caveat is that sexually antagonistic genes might increase the fitness of sons and reduce that of daughters. However, mating treatment did not differentially affect the survival of sons and daughters ($\chi^2 = 0.26$, $P = 0.61$; see Supplementary Information).

It has been argued, but not yet shown¹⁰, that polyandry confers genetic benefits because paternity is biased towards sires that generally produce more viable offspring ('good sperm' hypothesis)⁹. In *A. stuartii*, the ability of a male to gain a high share of paternity with three different polyandrous females was positively related to the survival of his offspring when he mated to a fourth female. Although we did not assign paternity at conception, we determined it ~6 d after embryo implantation. This reduces the likelihood that differential

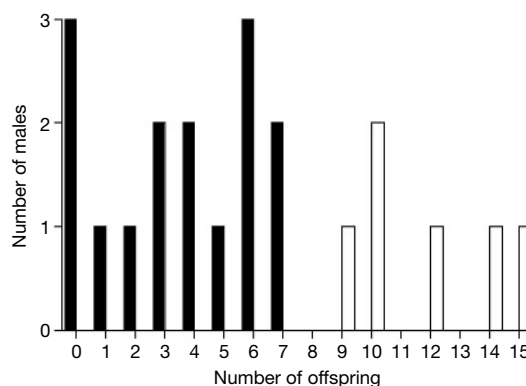


Figure 3 | Total number of offspring sired by males when mated to three polyandrous females. In the 2004 captive female experiment, males with 0–7 offspring were assigned to the 'low ejaculate competitiveness' category (filled bars) and those with 9–15 offspring to the 'high competitiveness' category (open bars).

mortality created a spurious relationship between offspring survival and estimated ejaculate competitiveness¹⁰. Moreover, there was negligible mortality in the first 35 d after birth, differential offspring mortality only occurred shortly before weaning, and sire ejaculate competitiveness was unrelated to litter size at birth. Because our estimates of the ability to gain paternity and the sire effect on offspring survival were independent, our results strongly support the 'good sperm' hypothesis.

In some species, polyandry reduces the production of inbred young^{2,8}, but this is unlikely in *A. stuartii*. Our experimental design precluded females from mating with close relatives (see Supplementary Information). However, our data do not shed light on whether polyandrous *A. stuartii* females can also bias paternity to reduce incompatibility between parental genotypes¹⁰. 'Genetic compatibility' and 'good sperm' explanations for polyandry are not mutually exclusive. The former requires non-additive genetic variation and the latter additive genetic variation for fitness²⁴. Both sources of variation can be significant simultaneously²⁵. Of course, both hypotheses also require that post-copulatory mechanisms can bias paternity in the appropriate direction^{1,4,5,10,24}. Our study shows that polyandry benefits female *A. stuartii* because males that sire fitter offspring gain more paternity under sperm competition. The strength of our results may be due in part to the unusual reproductive biology of this species. In *A. stuartii*, total spermatogenic failure occurs about 1 month before mating, and sperm are continually lost in the urine through spermatorrhoea. Male antechinus copulate for 5–14 consecutive hours with each female, and ejaculate around 3 h after mating starts^{18,21}. This extraordinary male reproductive biology could subject sperm to extreme physiological and epigenetic stress, resulting in the marked relationship between male sperm competitive ability and offspring viability (see Supplementary Information).

METHODS

Mating. We caught antechinus at Kioloa, Australia (35° 32' S, 150° 22' E) in August 2003 and 2004, shortly before the predictable onset of mating and transferred them to the laboratory²¹. We detected oestrus by daily monitoring of urine for cornified epithelial cells²³. Receptive females were alternately assigned to the two mating treatments to standardize the date of mating. In 2003, the 19 females in the monandrous treatment each mated three times to a uniquely assigned male and the 17 females in the polyandrous treatment each mated to three different males. We assigned mates from a pool of 41 males so that the cumulative number of prior matings did not differ between treatments. Females mated every second day. In 2004, we used a blocked experimental design. Each block had six females and three males. The three polyandrous females each mated once to each male such that each male mated once as the first, second and third mate (that is, ABC, BCA and CAB). The three monandrous females each mated three times to a single male (that is, AAA, BBB and CCC). Males therefore mated six times. Again, the cumulative number of prior matings per male did not differ between treatments. No antechinus mated more than once a day. After each mating, we confirmed that insemination was successful by examining female urine or a copulatory plug for the presence of sperm. This design provided independent estimates of the ability of a male to gain paternity and his effect on offspring survival.

Offspring paternity and survival. Females were housed individually and fed *ad libitum*²³. We checked pouches daily for young from 27 d after the first mating. In 2003, we measured the crown–rump length of each 32-d-old offspring, gave it a toe-bud clip and then, at 34 d old, released the family into a nestbox at the point of original capture. To calculate survival to weaning, we directly counted young in the nestbox when they were ~80–85 d old ($n = 12$ families) or, if the family had moved to a natural cavity nearby ($n = 36$), we intensively trapped outside the nest to catch young making initial exploratory forays²⁰. After weaning at ~85–111 d old, we also comprehensively trapped the site every fourth week until the following breeding season²⁰. Recaptured offspring were individually micro-chipped. In 2004, females were checked every 3 h (7 h overnight) from 27 d after their first mating to collect any offspring that failed to attach for genotyping. All young were genotyped (see Methods in refs 23, 26, and Supplementary Information). We sexed, individually marked and measured offspring as soon as they voluntarily detached from the teat. Offspring survival was monitored daily until 80–85 d of age, whereupon families were released at their site of capture. Our estimate of survival to weaning was therefore based on slightly

different criteria in 2004 (alive at ~85 d) and 2003 (recaptured shortly after weaning: ≥ 90 d old).

Statistical analysis. Analyses were run in R 2.3.1 (refs 27, 28) or Genstat 8.0 (ref. 29). The proportion of females with young on every teat was analysed with logistic regression (binomial error) with year (that is rearing environment) and either mating treatment or ejaculate competitiveness as factors. The effect of mating treatment on offspring survival to weaning was analysed as the proportion of a litter that survived (numerator, number alive; denominator, young on teats after birth) using generalized linear mixed models with binomial error (fixed factors: year, mating treatment; random factor: block in 2004). One block was excluded because a male seemed to be infertile and a female died. Separate analyses were run for each year. We also compared offspring survival of monandrous females mated to males of high and low ejaculate competitiveness and polyandrous females using the same modelling approach. We used monthly recapture data to generate Jolly–Seber estimates of post-weaning survival probabilities³⁰.

Received 2 July; accepted 6 September 2006.

- Jennions, M. D. & Petrie, M. Why do females mate multiply? A review of the genetic benefits. *Biol. Rev. Camb. Phil. Soc.* **75**, 21–64 (2000).
- Madsen, T., Shine, R., Loman, J. & Hakansson, T. Why do female adders copulate so frequently? *Nature* **335**, 440–441 (1992).
- Tregenza, T. & Wedell, N. Benefits of multiple mates in the cricket *Gryllus bimaculatus*. *Evolution Int. J. Org. Evolution* **52**, 1726–1730 (1998).
- Tregenza, T. & Wedell, N. Genetic compatibility, mate choice and patterns of parentage. *Mol. Ecol.* **9**, 1013–1027 (2000).
- Zeh, J. A. & Zeh, D. W. Reproductive mode and the genetic benefits of polyandry. *Anim. Behav.* **61**, 1051–1063 (2001).
- Olsson, M., Shine, R., Madsen, T., Gullberg, A. & Tegelstrom, H. Sperm selection by females. *Nature* **383**, 585 (1996).
- Thuman, K. A. & Griffith, S. C. Genetic similarity and the non-random distribution of paternity in a genetically highly polyandrous shorebird. *Anim. Behav.* **69**, 765–770 (2005).
- Tregenza, T. & Wedell, N. Polyandrous females avoid costs of inbreeding. *Nature* **415**, 71–73 (2002).
- Yasui, Y. A 'good sperm' model can explain the evolution of costly multiple mating by females. *Am. Nat.* **149**, 573–584 (1997).
- Simmons, L. W. The evolution of polyandry: sperm competition, sperm selection, and offspring viability. *Annu. Rev. Ecol. Syst.* **36**, 125–146 (2005).
- Tregenza, T., Wedell, N., Hosken, D. J. & Ward, P. I. Maternal effects on offspring depend on female mating pattern and offspring environment in yellow dung flies. *Evolution Int. J. Org. Evolution* **57**, 297–304 (2003).
- Sheldon, B. C. Differential allocation: tests, mechanisms and implications. *Trends Ecol. Evol.* **15**, 397–402 (2000).
- Kozłowska, M., Krzeminska, A. & Radwan, J. Good genes and the maternal effects of polyandry on offspring reproductive success in the bulb mite. *Proc. R. Soc. Lond. B* **271**, 165–170 (2004).
- Weigensberg, I., Carrière, Y. & Roff, D. A. Effects of male genetic contribution and paternal investment to egg and hatching size in the cricket *Gryllus firmus*. *J. Evol. Biol.* **11**, 135–146 (1998).
- Zeh, J. A. & Zeh, D. W. Outbred embryos rescue inbred half-siblings in mixed-paternity broods of live-bearing females. *Nature* **439**, 201–203 (2006).
- Evans, J. P., Zane, L., Francescato, S. & Pilastro, A. Directional postcopulatory sexual selection revealed by artificial insemination. *Nature* **421**, 360–363 (2003).
- Hosken, D. J., Garner, T. W. J., Tregenza, T., Wedell, N. & Ward, P. I. Superior sperm competitors sire higher-quality young. *Proc. R. Soc. Lond. B* **2270**, 1933–1938 (2003).
- Holleley, C. E., Dickman, C. R., Crowther, M. S. & Oldroyd, B. P. Size breeds success: multiple paternity, multivariate selection and male semelparity in a small marsupial, *Antechinus stuartii*. *Mol. Ecol.* **15**, 3439–3448 (2006).
- Selwood, L. A timetable of embryonic development of the dasyurid marsupial *Antechinus stuartii* (Macleay). *Aust. J. Zool.* **28**, 649–668 (1980).
- Fisher, D. O. Population density and presence of the mother are related to natal dispersal decisions in male and female *Antechinus stuartii*. *Aust. J. Zool.* **53**, 103–110 (2005).
- Fisher, D. O. & Cockburn, A. The large male advantage in brown antechinus: female choice, male dominance and delayed male death. *Behav. Ecol.* **17**, 164–171 (2005).
- Kraaijeveld-Smit, F. J. L., Ward, S. J. & Temple-Smith, P. D. Paternity success and the direction of sexual selection in a field population of a semelparous marsupial, *Antechinus agilis*. *Mol. Ecol.* **12**, 475–484 (2003).
- Fisher, D. O., Double, M. C. & Moore, B. D. Number of mates and timing of mating affect offspring growth in the small marsupial, *Antechinus agilis*. *Anim. Behav.* **71**, 289–297 (2006).
- Neff, B. D. & Pitcher, T. E. Genetic quality and sexual selection: an integrated framework for good genes and compatible genes. *Mol. Ecol.* **14**, 19–38 (2005).

25. Evans, J. P. & Marshall, D. J. Male-by-female interactions influence fertilization success and mediate the benefits of polyandry in the sea urchin *Heliocidaris erythrogramma*. *Evolution Int. J. Org. Evolution* **59**, 106–112 (2005).
26. Nicholls, J. A., Double, M. C., Rowell, D. & Magrath, R. D. The evolution of cooperative and pair breeding in thornbills *Acanthiza* (Aves: Pardalotidae). *J. Avian Biol.* **31**, 165–176 (2000).
27. Bates, D. & Maechler, M. Matrix: a matrix package for R (R Foundation for Statistical Computing, Vienna, 2006).
28. R Core Development Team. R: a language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, 2006).
29. Genstat Committee. *Genstat Manual, version 8* (Clarendon, Oxford, UK, 2005).
30. Pollock, K. H., Nichols, J. D., Brownie, C. & Hines, J. E. Statistical inference for capture-recapture experiments. *Wildlife Monogr.* **107**, 1–97 (1990).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Backwell, S. Griffith and M. Olsson for discussions and comments; numerous field and lab assistants for their hard work; P. Marsack, S. and R. Berkhout, and O. and C. Carriage for help at Kioloa. This work was supported by an Australian Research Council fellowship (to D.O.F.).

Author Contributions D.O.F. designed the main study, carried out the experiments, performed animal husbandry and drafted the manuscript. M.C.D. performed the molecular analysis. S.P.B. performed statistical analyses and contributed to fieldwork. A.C. carried out additional statistical analyses; M.D.J. designed the sperm competition experiment. All authors discussed the results and analysis and contributed to the writing.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.O.F. (diana.fisher@anu.edu.au).

Global distribution and conservation of rare and threatened vertebrates

Richard Grenyer^{1*†}, C. David L. Orme^{2*}, Sarah F. Jackson³, Gavin H. Thomas^{4†}, Richard G. Davies³, T. Jonathan Davies^{1†}, Kate E. Jones⁵, Valerie A. Olson^{5†}, Robert S. Ridgely⁶, Pamela C. Rasmussen⁷, Tzung-Su Ding⁸, Peter M. Bennett⁵, Tim M. Blackburn⁴, Kevin J. Gaston³, John L. Gittleman^{1†} & Ian P. F. Owens^{2,9}

Global conservation strategies commonly assume that different taxonomic groups show congruent geographical patterns of diversity, and that the distribution of extinction-prone species in one group can therefore act as a surrogate for vulnerable species in other groups when conservation decisions are being made^{1–4}. The validity of these assumptions remains unclear, however, because previous tests have been limited in both geographical and taxonomic extent^{5–12}. Here we use a database on the global distribution of 19,349 living bird, mammal and amphibian species to show that, although the distribution of overall species richness is very similar among these groups, congruence in the distribution of rare and threatened species is markedly lower. Congruence is especially low among the very rarest species. Cross-taxon congruence is also highly scale dependent, being particularly low at the finer spatial resolutions relevant to real protected areas. ‘Hotspots’ of rarity and threat are therefore largely non-overlapping across groups, as are areas chosen to maximize species complementarity. Overall, our results indicate that ‘silver-bullet’ conservation strategies alone will not deliver efficient conservation solutions. Instead, priority areas for biodiversity conservation must be based on high-resolution data from multiple taxa.

Our analyses are based on three high-resolution databases of the global distribution of birds, mammals and amphibians^{13–15}. For each group, we mapped the geographical distribution of species richness for all species, rare species and threatened species (Fig. 1). These maps were based on grid cells roughly equivalent to 1° latitude by 1° longitude. For each aspect of richness, we calculated the pair-wise Pearson’s correlation coefficient (r) for grid-cell richness values of the three vertebrate classes, controlling for spatial covariance when estimating the statistical significance of correlations¹⁶ (Fig. 2a). Although all pair-wise correlations were positive and statistically significant at a global scale, the cross-taxon congruence varied markedly across the three aspects of richness. There was very high congruence with respect to total species richness ($0.79 \leq r \leq 0.90$), but congruence among rare species was lower ($0.24 \leq r \leq 0.48$), and lower still among threatened species ($0.13 \leq r \leq 0.32$). Typically, mammals and birds showed the highest global congruence, whereas amphibians showed the lowest. Caution needs to be exercised when invoking explanations for these differences, however, because amphibians have substantially smaller ranges (median range $5.6 \times 10^4 \text{ km}^2$) than mammals ($5.8 \times 10^5 \text{ km}^2$) or birds ($1.0 \times 10^6 \text{ km}^2$). The low overlap between amphibian ranges and

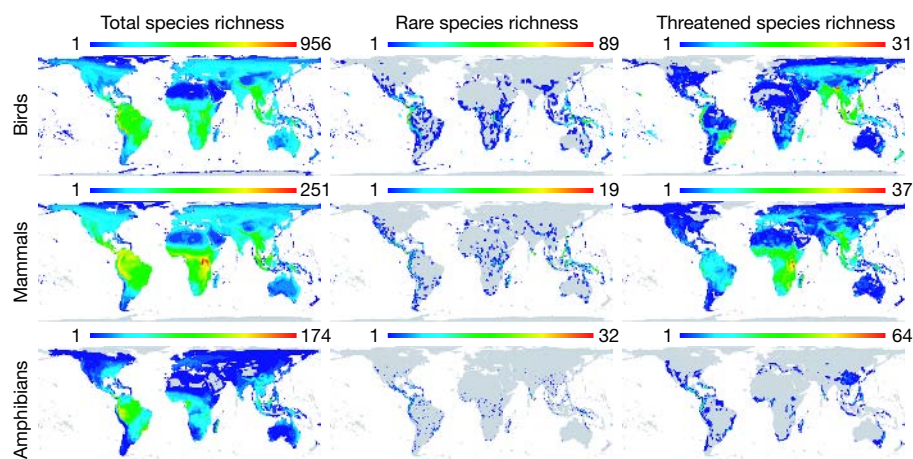


Figure 1 | Global richness maps for birds, mammals and amphibians. Global richness is shown with respect to the three different aspects of species richness used here. Colour gradients are linear with respect to species number.

¹Department of Biology, Gilmer Hall, University of Virginia, Charlottesville, Virginia 22904, USA. ²Division of Biology, Imperial College London, Silwood Park, Ascot, Berkshire SL5 7PY, UK. ³Biodiversity & Macroecology Group, Department of Animal & Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK. ⁴School of Biosciences, University of Birmingham, Birmingham B15 2TT, UK. ⁵Institute of Zoology, Zoological Society of London, London NW1 4RY, UK. ⁶Academy of Natural Sciences, Philadelphia, Pennsylvania 19103, USA. ⁷Michigan State University Museum and Department of Zoology, Michigan 48824-1045, USA. ⁸School of Forestry & Resource Conservation, National Taiwan University, Taipei 106, Taiwan. ⁹Natural Environment Research Council Centre for Population Biology, Silwood Park, Ascot, Berkshire, SL5 7PY, UK. †Present addresses: Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, UK (R.G.); NERC Centre for Population Biology, Imperial College London, Silwood Park, Ascot, Berkshire SL5 7PY, UK (G.H.T.); Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK (V.A.O.); Institute of Ecology, University of Georgia, Athens, Georgia 30602, USA (J.L.G., T.J.D.).

*These authors contributed equally to this work.

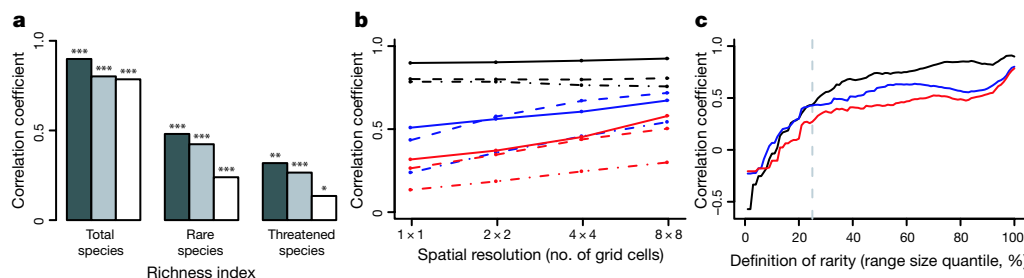


Figure 2 | Cross-taxon congruence and the effects of scale and definition of rarity on congruence. **a**, Congruence for overall species richness, rare species and threatened species between birds and mammals (dark grey bars), birds and amphibians (light grey bars) and mammals and amphibians (open bars). Asterisks indicate statistical significance controlling for spatial non-independence (* $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$). **b**, Effect of scale on congruence for overall species richness (black), rare species (blue) and

threatened species (red). Congruence between birds and mammals (unbroken lines), birds and amphibians (dashed lines) and mammals and amphibians (dotted-dashed lines). **c**, Relationship between definition of rarity and congruence for rare species of birds and mammals (black), birds and amphibians (blue) and mammals and amphibians (red). Definition of rarity refers to the quantile of range distribution. The dashed, vertical line indicates the lower quartile of species.

those of the other two groups may, therefore, simply reflect the null expectation that small ranges are less likely to overlap with other ranges. In general, these global patterns of congruence held within biogeographic realms and biomes¹⁷ (Supplementary Fig. 1), with congruence being particularly high in the tropics and particularly low across the Holarctic. This latter finding may reflect differences between groups in their ability to re-colonize areas after glacial retreat.

Congruence among rare and threatened species varied with the resolution of data used in the analysis¹⁸ and increased markedly at coarser resolutions (Fig. 2b). This observation may explain differences in results between our study and those using coarser scale data^{2,4,12}. An analysis using ecoregions¹⁷, for example, reported high cross-taxon congruence for rare ('endemic') species ($0.49 < r < 0.61$)¹². The average size of those ecoregions (median area $5.5 \times 10^4 \text{ km}^2$), however, is much larger than our grid cells and is several thousand times larger than most protected areas (1.53 km^2 ; ref. 19 and Supplementary Fig. 2). Our results show that congruence among rare and threatened species declines rapidly as the scale approaches that more relevant to real protected areas. High congruence at the ecoregion scale does

not, therefore, mean that reserves in ecoregions will also show high congruence.

Congruence among rare species also varied with the definition of rarity, and observed congruence decreased as the definition became more stringent (Fig. 2c). Indeed, across the very rarest species in each class (for example, those with the smallest 10% of ranges), all pair-wise correlations between groups were negative: the very rarest birds, mammals and amphibians inhabit different places from one another. Such patterns may explain why previous analyses have reported high congruence for rare species^{2,4,12}. Although our definitions of rare species correspond to ranges of $<278,250 \text{ km}^2$ in birds, $<41,685 \text{ km}^2$ in mammals and $<430 \text{ km}^2$ in amphibians, previous analyses defined rare ('endemic') species simply as those recorded as occupying single hotspots ($<2,373,057 \text{ km}^2$; ref. 4) or ecoregions ($<4,629,589 \text{ km}^2$; ref. 17). Our results show that such relaxed definitions of rarity lead to raised estimates of congruence.

We also tested whether geographical patterns of richness in one group act as a surrogate for those in other groups^{6,7,9,11}. We tested whether richness hotspots (the richest 5% of grid cells) identified for one group overlapped with corresponding hotspots for other groups. Hotspots of total species richness showed high congruence, with 17.8% of all hotspot grid cells common across groups (Fig. 3a). Hotspots of rare or threatened species showed much lower congruence, however, with only 2.3% of rarity hotspot cells (Fig. 3b) and 0.6% of threat hotspot cells (Fig. 3c) common across groups. These patterns remained intact when we used an optimal complementarity algorithm to choose sets of grid cells that efficiently capture all species in a group (Table 1), and even when we controlled for differences in the area of the sets (Supplementary Table 1). We measured surrogacy as the proportion of species from other groups represented within sets²⁰.

Finally, we compared the performance of different methods of identifying priority areas with respect to their ability to capture multi-taxon diversity. For overall species richness, there were no substantial differences between sets based on single taxa and those based on multiple taxa, in terms of either area covered or species captured (Fig. 4a). For rare and threatened species, however, sets based on multiple taxa captured far more species than did single taxon sets in a similar overall area (Fig. 4b, c, and Supplementary Fig. 3). Complementarity sets based on high-resolution data on multiple taxa were also much more efficient in capturing rare and threatened species than were areas identified in previous global analyses^{1,4,21}, even when the complementarity sets represented each species more than once²⁰ (Fig. 4, and Supplementary Fig. 3). Indeed, areas identified in some previous global analyses often did not perform any better than randomly selected sets of cells of equivalent overall size, and in some cases performed worse (Fig. 4).

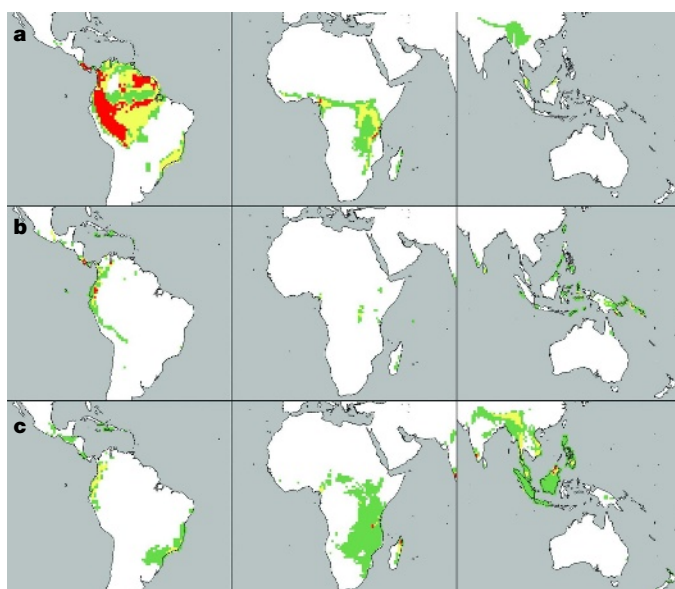


Figure 3 | Cross-taxon congruence of richness hotspots. Hotspots of total species (a), rare species (b) and threatened species (c) richness. Red shading shows cells that are hotspots for all three groups, yellow for two groups, and green for one group. Hotspots are the richest 5% of non-zero cells.

Our finding that cross-taxon congruence is high for total species richness, given that the richest areas are consistently associated with low latitudes and mountainous regions¹³, reflects the importance of the interaction between energy and topography in predicting diversity. Caution is again needed when invoking explanations for why congruence and surrogacy are lower for rare and threatened species: such species typically have relatively small geographic ranges and thus low overlap in range might be expected. There may, however, be additional factors. For rare species, all three groups have richness peaks on the neotropical mainland, but avian rarity also peaks on oceanic island archipelagos, whereas rare mammal species are concentrated on continental shelf islands and rare amphibian species are often on continental land masses. Such differences may therefore reflect relative dispersal ability. For threatened species, low congruence may also result from differences among groups in their sensitivity to particular threatening processes. Although the main source of threat for all three groups is habitat loss, subsidiary threats differ among groups. Invasive species and overexploitation are chief secondary sources in birds³, overexploitation is the main secondary source in mammals³, and climate change, pollution and transmissible disease are important in amphibians¹⁵. Low congruence among threatened species may therefore be driven by differences in the distribution of these risks.

Could our findings that cross-taxon congruence is high for overall species richness but low for rare and threatened species be due to biases in the databases? Systematic error could result if some geographical regions, or some taxa, are relatively poorly studied; however, our key findings remain qualitatively intact even if we restrict our analyses within the best-studied areas (the Nearctic, Palearctic and Australia; Supplementary Fig. 1) or the best known groups (birds and mammals; Table 1). Bias in our estimates of overall range could also influence cross-taxon congruence because congruence should decline with decreasing range size. Our conclusions are probably conservative in this respect, however, because the extents of occurrence used tend to overestimate true ranges²². It is therefore unlikely that our main conclusions are artefactual: indeed, we predict that

Table 1 | Patterns of cross-taxon surrogacy across birds, mammals and amphibians

Richness index	Surrogate groups	$N_{\text{spp.}}$	N_{cells}	Target groups* (% of species represented in set)		
				Birds	Mammals	Amphibians
Total	Birds	9,626	421	–	79.4 ± 0.3	55.5 ± 0.7
	Mammals	4,104	509	91.7 ± 0.1	–	61.1 ± 0.5
	Amphibians	5,619	831	90.9 ± 0.2	86.2 ± 0.2	–
	Birds, mammals	13,730	714	–	–	68.4 ± 0.5
	Birds, amphibians	15,245	1,028	–	89.8 ± 0.2	–
	Mammals, amphibians	9,723	1,077	95.0 ± 0.1	–	–
	All three groups	19,349	1,223	–	–	–
Rarity	Birds	2,424	380	–	43.3 ± 1.1	22.5 ± 1.3
	Mammals	1,026	432	68.3 ± 0.4	–	27.0 ± 0.7
	Amphibians	1,405	560	63.7 ± 0.5	51.6 ± 0.6	–
	Birds, mammals	3,450	656	–	–	35.7 ± 0.9
	Birds, amphibians	3,829	808	–	63.1 ± 0.6	–
	Mammals, amphibians	2,431	858	77.9 ± 0.2	–	–
	All three groups	4,855	1,033	–	–	–
Threat	Birds	1,096	282	–	51.7 ± 0.9	31.2 ± 1.4
	Mammals	1,033	357	60.7 ± 0.6	–	39.7 ± 0.9
	Amphibians	1,856	454	62.7 ± 0.4	59.7 ± 0.4	–
	Birds, mammals	2,129	518	–	–	49.2 ± 0.6
	Birds, amphibians	2,952	627	–	67.2 ± 0.5	–
	Mammals, amphibians	2,889	690	72.4 ± 0.4	–	–
	All three groups	3,985	821	–	–	–

* Values are the percentage of species in the target groups represented in complementary sets of grid cells designed to contain all members of the surrogate groups (mean ± s.d. over 100 such sets). $N_{\text{spp.}}$ is total number of species in the surrogate group. N_{cells} is number of cells in the optimal complementarity set.

more detailed information on the geographical distribution of poorly known species will show that cross-taxon congruence and surrogacy for rare and threatened species are even lower than estimated here.

Our findings should be interpreted cautiously with respect to applied conservation. We have ignored the political and socio-economic factors that are vital in practical conservation^{23,24} and have focused on species rather than ecoregions and ecosystem services^{17,25}. We have also restricted our analyses to terrestrial habitats and closely related organisms. Nevertheless, we have shown that, even among terrestrial vertebrates, the extent to which rare and threatened species from one group can act as a surrogate for corresponding species in other groups is severely limited, especially at the finer scales most relevant to conservation. At such scales, we predict that low cross-taxon congruence will be a property of any set of global priority areas and that congruence is likely to be even lower among more distantly related organisms or across terrestrial and aquatic habitats. This is of concern because, although previous global analyses have explored numerous methods for identifying priority areas²⁵, they are fundamentally based on data from just one or two groups^{1,2,4,21}. Our results suggest that designing effective protected area networks will require high-resolution data on the distribution of multiple taxa and an understanding of how these relate to ecosystems. These challenges are being tackled through projects mapping the distribution of species at a scale comparable to individual protected areas^{26–28}. We anticipate, however, that ‘silver-bullet’ conservation strategies based on particular taxonomic groups will not be effective because locations rich in one aspect of diversity will not necessarily be rich in others.

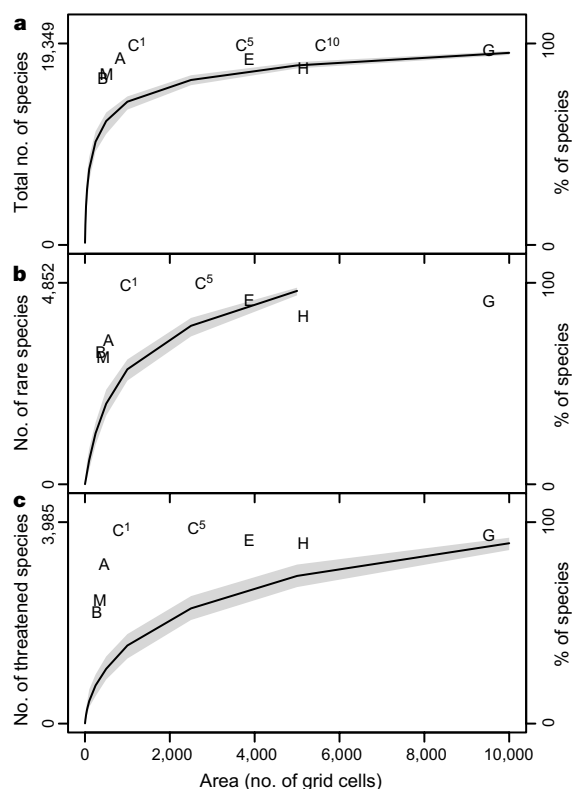


Figure 4 | Relative performance of different types of priority network. Shown is performance with respect to capturing total species (a), rare species (b) and threatened species (c) richness of birds, mammals and amphibians. Networks were identified by using an optimal complementarity approach based on birds alone (B), mammals alone (M), amphibians alone (A), or birds, mammals and amphibians combined (C^n , where n indicates the number of times each species is represented). Performance is also shown for biodiversity hotspots (H), endemic bird areas¹ (E) and the global 200 ecoregions²¹ (G), and for randomly selected sets of cells (100 replicates: median, black line; 95% confidence range, grey area).

METHODS

Databases. Analyses were based on vector range maps of 9,626 species of terrestrial birds¹³, 4,104 species of terrestrial mammals¹⁴, and 5,619 species of amphibians²⁷. Range maps were projected onto a Behrmann equal-area projection and converted to a grid with resolution $96.3 \times 96.3 \text{ km}^2$ (ref. 13). Species richness was the total number of species present in each grid cell¹³. Rare species richness was the total number of rare species present, where rare species were those in the lower quartile of the range distribution of each taxonomic group^{13,29}. A relative definition of rarity was used rather than an absolute one because of the large difference in absolute geographic range between the three vertebrate classes (median range: birds, $1,001,559 \text{ km}^2$; mammals, $574,969 \text{ km}^2$; amphibians, $55,642 \text{ km}^2$). No definition of rarity based on an absolute range area therefore successfully identified a consistent proportion of species in each taxonomic class. Threatened species richness was the total number of the threatened species present¹³, where threatened species were those classified by The World Conservation Union (IUCN) as 'vulnerable', 'endangered' or 'critically endangered'³⁰. Biogeographic realms and biomes were those identified in ref. 17. Median ranges for each group were based on grid-cell counts.

Cross-taxon congruence. Cross-taxon congruence was measured with Pearson correlation coefficients among grid cells, calculated for all pair-wise combinations of the three vertebrate classes for each of total, rare and threatened species richness. To control for spatial non-independence, statistical significance was calculated under an estimated effective sample size given the observed degree of spatial autocorrelation¹⁶. We adjusted sample size, rather than explicitly modelling spatial non-independence¹³, because at this scale the cumulative number of pair-wise correlations required across all tests is very large and thus would be prohibitive to run with spatially explicit models.

Scale of analysis and definition of rarity. Analyses were conducted at four resolutions: 1×1 grid cell ($\sim 9,274 \text{ km}^2$), 2×2 grid cells ($\sim 37,000 \text{ km}^2$), 4×4 grid cells ($\sim 150,000 \text{ km}^2$) and 8×8 grid cells ($\sim 600,000 \text{ km}^2$). The criterion for rarity varied in stringency from rare species being those in the first percentile of the range distribution for their class (highest stringency) to rare species being those in the lower 99 percentiles of that distribution (lowest stringency).

Hotspots. Hotspots were defined as the richest 5% of grid cells¹³ for each richness index, and overlap was quantified as the number of hotspot grid cells common to all three taxonomic groups as a percentage of the total number of hotspot cells^{5,13}. Varying the percentage used to define hotspots¹³ did not qualitatively change the pattern of overlap.

Optimal complementarity sets. We calculated the smallest number of grid cells required to represent each species in the 'surrogate' taxonomic group at least once and then identified 100 such sets of cells²⁰. The surrogacy value of a group was calculated by counting the average number of species from the other 'target' groups represented in each set'. To control for differences in set size when comparing surrogacy, we selected the richest 250 cells (the approximate size of the smallest single-taxon set) from each set. Complementarity sets were also calculated with more than one taxon in the 'surrogate' group and by using an algorithm that attempted to represent each species in the surrogate set more than once²⁰. Performance of complementarity sets was measured as the number of species that the sets contained; this value was compared with that of randomly selected sets of cells, and with areas identified in previous global analyses^{1,4,21}, which were rasterized to our grid. Species composition of cells was based on our databases and our definitions of rarity and threat.

Received 24 July; accepted 8 September 2006.

1. Stattersfield, A. J., Crosby, M. J., Long, A. J. & Wege, D. C. *Endemic Bird Areas of the World: Priorities for Biodiversity Conservation* (BirdLife International, Cambridge, UK, 1998).
2. Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
3. Baillie, J. E. M., Hilton-Taylor, C. & Stuart, S. N. *2004 IUCN Red List of Threatened Species: A Global Species Assessment* (IUCN, Gland, Switzerland, and Cambridge, UK, 2004).
4. Mittermeier, R. A. *et al.* *Hotspots Revisited: Earth's Biologically Richest and Most Endangered Terrestrial Ecoregions* (Cemex, Conservation International and Agrupación Sierra Madre, Monterrey, Mexico, 2005).
5. Prendergast, J. R., Quinn, R. M., Lawton, J. H., Eversham, B. C. & Gibbons, D. W. Rare species, the coincidence of diversity hotspots and conservation strategies. *Nature* **365**, 335–337 (1993).
6. Dobson, A. P., Rodriguez, J. P., Roberts, M. W. & Wilcove, D. S. Geographic distribution of endangered species in the United States. *Science* **275**, 550–553 (1993).

7. Howard, P. C. *et al.* Complementarity and the use of indicator groups for reserve selection in Uganda. *Nature* **394**, 472–475 (1998).
8. Lawton, J. H. *et al.* Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. *Nature* **391**, 72–76 (1998).
9. van Jaarsveld, A. S. *et al.* Biodiversity assessment and conservation strategies. *Science* **279**, 2106–2108 (1998).
10. Brooks, T. *et al.* Toward a blueprint for conservation in Africa. *Bioscience* **51**, 613–624 (2001).
11. Moore, J. L. *et al.* Performance of sub-Saharan vertebrates as indicator groups for identifying priority areas for conservation. *Conserv. Biol.* **17**, 207–218 (2003).
12. Lamoreux, J. F. *et al.* Global tests of biodiversity concordance and the importance of endemism. *Nature* **440**, 212–214 (2006).
13. Orme, C. D. L. *et al.* Global hotspots of species richness are not congruent with endemism or threat. *Nature* **436**, 1016–1019 (2005).
14. Sechrest, W. W. *Global Diversity, Endemism, and Conservation of Mammals*. Thesis, Univ. Virginia (2003).
15. Stuart, S. N. *et al.* Status and trends of amphibian declines and extinctions worldwide. *Science* **306**, 1783–1786 (2004).
16. Clifford, P., Richardson, S. & Hemon, D. Assessing the significance of the correlation between two spatial processes. *Biometrics* **45**, 123–134 (1989).
17. Olson, D. M. *et al.* Terrestrial ecoregions of the world: a new map of life on Earth. *Bioscience* **51**, 933–938 (2001).
18. Pearson, D. L. & Carroll, S. S. The influence of spatial scale on cross-taxon congruence patterns and prediction accuracy of species richness. *J. Biogeogr.* **26**, 1079–1090 (1999).
19. WDP Consortium. 2005 World Database on Protected Areas (http://gis.tnc.org/data/IMS/WDP_viewer/WDP_info/index.html) (2005).
20. Rodrigues, A. S. L., Cerdeira, J. & Gaston, K. J. Flexibility, efficiency and accountability: adapting reserve selection algorithms to more complex conservation problems. *Ecography* **23**, 565–574 (2000).
21. Olson, D. M. & Dinerstein, E. The Global 200: priority ecoregions for global conservation. *Ann. Missouri Bot. Gard.* **89**, 199–224 (2002).
22. Orme, C. D. L. *et al.* Global patterns of geographic range size in birds. *PLoS Biol.* **4**, e208 (2006).
23. Pressey, R. L., Humphries, C. J., Margules, C. R., Vane-Wright, R. I. & Williams, P. H. Beyond opportunism: key principles for systematic reserve selection. *Trends Ecol. Evol.* **8**, 124–128 (1993).
24. Wilson, K. A., McBride, M. F., Bode, M. & Possingham, H. P. Prioritizing global conservation efforts. *Nature* **440**, 337–340 (2006).
25. Brooks, T. M. *et al.* Global biodiversity conservation priorities. *Science* **313**, 58–61 (2006).
26. Birdlife International. World Bird Database (<http://www.birdlife.org/datazone/index.html>) (2006).
27. IUCN, Conservation International & NatureServe. Global Amphibian Assessment (<http://www.globalamphibians.org>) (2004).
28. IUCN, Conservation International & NatureServe. Global Mammal Assessment (http://www.iucn.org/themes/ssc/biodiversity_assessments/gma/indexgma.htm) (2006).
29. Gaston, K. J. *Rarity* (Chapman & Hall, London, 1994).
30. IUCN. 2006 IUCN Red List of Threatened Species (<http://www.iucnredlist.org>) (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the field biologists who have made these data sets possible; the Global Amphibian Assessment, BirdLife International, Conservation International, IUCN, NatureServe, A. Stattersfield, S. Stuart, W. Sechrest and L. Boitani for access to data; J. Baillie, M. Balman, L. Bennun, L. Boitani, T. Brooks, M. Burgess, S. Butchart, M. Cardillo, F. Eigenbrod, S. Fritz, C. Godfray, G. Mace, S. Meiri, J. O'Dell, A. Phillimore, N. Pickup, A. Purvis, W. Sechrest, E. Smith, A. Stattersfield, S. Stuart, A. Webster and the London e-Science Centre for technical assistance, discussion or comments on the manuscript. This work was funded by the National Science Foundation and the UK Natural Environment Research Council. K.J.G. holds a Royal Society–Wolfson Research Merit Award.

Author Contributions Team leaders J.L.G. and I.P.F.O. contributed equally to this work. The study was devised by R.G., C.D.L.O., J.L.G. and I.P.F.O. R.G., C.D.L.O. and S.F.J. conducted the analyses. R.G., C.D.L.O., S.F.J., G.H.T., R.G.D., T.J.D., V.A.O., R.S.R., P.C.R., T.-S.D., P.M.B., T.M.B., K.J.G., J.L.G. and I.P.F.O. contributed data or technical expertise. R.G., C.D.L.O., J.L.G. and I.P.F.O. wrote the initial draft of the manuscript. All authors commented on subsequent drafts.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.G. (r.grenyer@kew.org), or I.P.F.O. (i.owens@imperial.ac.uk) or J.L.G. (jl@uga.edu) for database enquiries.

Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*

Jörg Kämper¹, Regine Kahmann¹, Michael Bölker², Li-Jun Ma³, Thomas Brefort¹, Barry J. Saville^{4†}, Flora Banuett⁵, James W. Kronstad⁶, Scott E. Gold⁷, Olaf Müller¹, Michael H. Perlin⁸, Han A. B. Wösten⁹, Ronald de Vries⁹, José Ruiz-Herrera¹⁰, Cristina G. Reynaga-Peña¹⁰, Karen Snetselaar¹¹, Michael McCann¹¹, José Pérez-Martín¹², Michael Feldbrügge¹, Christoph W. Basse¹, Gero Steinberg¹, Jose I. Ibeas¹³, William Holloman¹⁴, Plinio Guzman¹⁵, Mark Farman¹⁶, Jason E. Stajich¹⁷, Rafael Sentandreu¹⁸, Juan M. González-Prieto¹⁹, John C. Kennell²⁰, Lazaro Molina¹, Jan Schirawski¹, Artemio Mendoza-Mendoza¹, Doris Greilinger¹, Karin Münch¹, Nicole Rössel¹, Mario Scherer¹, Miroslav Vraneš¹, Oliver Ladendorf¹, Volker Vincon¹, Uta Fuchs¹, Björn Sandrock², Shaowu Meng⁴, Eric C. H. Ho⁴, Matt J. Cahill⁴, Kylie J. Boyce⁶, Jana Klose⁶, Steven J. Klosterman⁷, Heine J. Deelstra⁹, Lucila Ortiz-Castellanos¹⁰, Weixi Li²¹, Patricia Sanchez-Alonso¹⁵, Peter H. Schreier²², Isolde Häuser-Hahn²², Martin Vaupel²², Edda Koopmann²², Gabi Friedrich²², Hartmut Voss²³, Thomas Schlüter²³, Jonathan Margolis²⁴, Darren Platt²⁴, Candace Swimmer²⁴, Andreas Gnirke²⁴, Feng Chen²⁴, Valentina Vysotskaia²⁴, Gertrud Mannhaupt^{1,25}, Ulrich Güldener²⁵, Martin Münsterkötter²⁵, Dirk Haase²⁵, Matthias Oesterheld²⁵, Hans-Werner Mewes^{25,26}, Evan W. Mauceli³, David DeCaprio³, Claire M. Wade³, Jonathan Butler³, Sarah Young³, David B. Jaffe³, Sarah Calvo³, Chad Nusbaum³, James Galagan³ & Bruce W. Birren³

Ustilago maydis is a ubiquitous pathogen of maize and a well-established model organism for the study of plant–microbe interactions¹. This basidiomycete fungus does not use aggressive virulence strategies to kill its host. *U. maydis* belongs to the group of biotrophic parasites (the smuts) that depend on living tissue for proliferation and development². Here we report the genome sequence for a member of this economically important group of biotrophic fungi. The 20.5-million-base *U. maydis* genome assembly contains 6,902 predicted protein-encoding genes and lacks pathogenicity signatures found in the genomes of aggressive pathogenic fungi, for example a battery of cell-wall-degrading enzymes. However, we detected unexpected genomic features responsible for the pathogenicity of this organism. Specifically, we found 12 clusters of genes encoding small secreted proteins with unknown function. A significant fraction of these genes exists in small gene families. Expression analysis showed that most of the genes contained in these clusters are regulated together and induced in infected tissue. Deletion of individual clusters altered the virulence of *U. maydis* in five cases, ranging from a complete

lack of symptoms to hypervirulence. Despite years of research into the mechanism of pathogenicity in *U. maydis*, no ‘true’ virulence factors³ had been previously identified. Thus, the discovery of the secreted protein gene clusters and the functional demonstration of their decisive role in the infection process illuminate previously unknown mechanisms of pathogenicity operating in biotrophic fungi. Genomic analysis is, similarly, likely to open up new avenues for the discovery of virulence determinants in other pathogens.

Ustilago maydis is a pathogenic basidiomycete fungus that infects maize, one of the world’s major cereal crops. The disease results in stunted plant growth and reduces yield, leading to severe economic losses¹. *U. maydis* is dimorphic (Fig. 1a) and grows in its haploid phase as a saprophytic yeast (Fig. 1c). Sexual development is initiated by the fusion of two haploid cells (Fig. 1d). The resulting dikaryon is filamentous and invades plant cells by means of a specialized infection structure called an appressorium (Fig. 1e). During penetration, the host plasma membrane invaginates and surrounds the invading hypha. An interaction zone develops between plant and fungal

¹Department of Organismic Interactions, Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Strasse, D-35043 Marburg, Germany. ²Department of Biology, Philipps-University Marburg, Karl-von-Frisch-Strasse 8, D-35032 Marburg, Germany. ³The Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. ⁴Department of Biology, University of Toronto at Mississauga, 3359 Mississauga Rd, Mississauga, Ontario, L5L 1C6, Canada. ⁵Department of Biological Sciences, California State University, 1250 Bellflower Boulevard, Long Beach, California 90840, USA. ⁶The Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, British Columbia, V6T 1Z4, Canada. ⁷Department of Plant Pathology, The University of Georgia, Athens, Georgia 30602-7274, USA. ⁸Department of Biology, Program on Disease Evolution, University of Louisville, Louisville, Kentucky 40292, USA. ⁹Department of Microbiology, Institute of Biomembranes, Utrecht University, Padualaan 8, 3705 SN Utrecht, The Netherlands. ¹⁰Departamento de Ingeniería Genética, Unidad de Biotecnología e Ingeniería Genética de Plantas, Centro de Investigación y de Estudios Avanzados del IPN, 36500 Irapuato, Gto. Mexico. ¹¹Department of Biology, Saint Joseph’s University, 5600 City Ave, Philadelphia, Pennsylvania 19131, USA. ¹²Department of Microbial Biotechnology, Centro Nacional de Biotecnología-CSIC, Campus de Cantoblanco-UAM, 28049 Madrid, Spain. ¹³Centro Andaluz de Biología del Desarrollo, CSIC/Universidad Pablo de Olavide, Carretera de Utrera Km1, 41013 Sevilla, Spain. ¹⁴Department of Microbiology and Immunology, Weill Medical College of Cornell University, 1300 York Avenue, New York, New York 10021, USA. ¹⁵Department of Plant Genetic Engineering, Cinvestav, Campus-Guanajuato, Km 9.6 Libramiento Norte, 36821 Irapuato, Guanajuato, Mexico. ¹⁶Department of Plant Pathology, University of Kentucky, 1405 Veteran’s Drive, Lexington, Kentucky 40546, USA. ¹⁷Department of Molecular Genetics and Microbiology, Duke University, Durham, North Carolina 27710, USA. ¹⁸Departamento de Microbiología i Ecología, Facultat de Farmàcia, Universitat de València, Vicent Andrés Estelles s/n, 46100 Burjassot, València, Spain. ¹⁹Departamento de Biotecnología Vegetal II, Centro de Biotecnología Genómica-IPN, Blvd del Maestro s/n, Cd. Reynosa, Tamaulipas, 88710, México. ²⁰Department of Biology, St Louis University, 3507 Laclede Avenue, St Louis, Missouri 63103-2010, USA. ²¹Department of Biology, University of Kentucky, 101 T. H. Morgan Building, Lexington, Kentucky 40546, USA. ²²Bayer CropScience AG, Alfred-Nobel-Strasse 50, D-40789 Monheim, Germany. ²³LION bioscience AG, Waldhofer Strasse 98, D-69123 Heidelberg, Germany. ²⁴Exelixis, Inc., 210 East Grand Avenue, South San Francisco, California 94080, USA. ²⁵Institute for Bioinformatics (MIPS), GSF National Research Center for Environment and Health, Ingolstädter Landstrasse 1, D-85764 Neuherberg, Germany. ²⁶Genome-oriented Bioinformatics, Technische Universität München, Am Forum 1, D-85354 Freising-Weihenstephan, Germany. †Present address: Forensic Science Program, Trent University, Peterborough, Ontario, K9J 7B8, Canada.

membranes that is characterized by fungal deposits produced by exocytosis⁴ (Fig. 1f). Although hyphae traverse plant cells, there is no apparent host defence response and plant tissue remains alive until late in the infection process. The most characteristic symptom of the disease is large tumours (Fig. 1b), which result from fungus-induced alterations in plant growth. The fungus proliferates and differentiates within tumour tissue (Fig. 1h) and produces masses of black diploid spores (Fig. 1g, i). On germination, spores undergo meiosis and produce the haploid phase⁵.

Genome analysis was performed on the haploid *U. maydis* strain 521 with the use of whole-genome shotgun (10× coverage) and map-based approaches (2.92× coverage; see Supplementary Methods). The assembly comprises 19.8 million bases (Mb) of the estimated genome size of 20.5 Mb (ref. 6). More than 99% of the assembly is represented in the 24 largest scaffolds (Supplementary Methods), which correspond to 23 identified chromosomes. Only chromosome 4 consists of two large scaffolds separated by the ribosomal DNA repeats (Supplementary Fig. S1). The final assembly contains 251 sequence gaps. Subtelomeric regions were identified for all except two chromosomes (Supplementary Table S1). Of more than 30,000

expressed sequence tags (ESTs; longer than 150 base pairs (bp); Supplementary Methods), 99.6% could be aligned with the genomic sequence, indicating almost complete sequence coverage. Independently, strain FB1—closely related to strain 521 by inbreeding⁷—was shotgun sequenced to 5× coverage, yielding an assembly of 19.3 Mb (a detailed description of the pedigree of strains FB1 and 521 is given in Supplementary Methods). A total of 18.9 Mb of this assembly can be aligned to the strain 521 assembly with 99.97% nucleotide sequence identity.

The *U. maydis* genome (20.5 Mb)⁶ is rather small in comparison with genomes of other plant pathogenic fungi (see the Broad Institute's Fungal Genome Initiative (FGI) Candidate Genome website, <http://www.broad.mit.edu/annotation/fungi/fgi/candidates.html>). The MIPS *Ustilago maydis* database (MUMDB; <http://mips.gsf.de/genre/proj/ustilago/>) currently lists 6,902 gene models, whereas in the phytopathogenic ascomycetes, gene numbers are considerably higher (12,841 in *Magnaporthe grisea*, 16,597 in *Stagonospora nodorum* and 11,640 in *Fusarium graminearum*; see the FGI website). The small number of genes is partly reflected in the absence of significant expansions of gene families (Supplementary Table S2). The small number of introns and their short mean length qualify as additional explanations for the small genome size of *U. maydis*. The average number of introns per gene is 0.46, with 70% of genes containing no intron. The related basidiomycetes *Cryptococcus neoformans*, *Coprinus cinereus* and *Phanerochaete chrysosporium* contain an average of 5.3, 4.5 and 2.6 introns per gene, respectively⁸. One outstanding example is the highly conserved *tor1* kinase gene, which lacks introns in *U. maydis* but contains 23 in *C. neoformans*. Apparently, the *U. maydis* genome has been shaped by massive, lineage-specific intron loss, as has been observed in the ascomycetous yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*⁹. Intron loss has been proposed to occur by the recombination of reverse-transcribed transcripts with the genomic copy⁹. The small number of introns observed in *U. maydis* might therefore be a consequence of the highly efficient homologous recombination system^{10,11}.

The genome of *U. maydis* does not show signs of large-scale duplication events (Supplementary Fig. S2) and is largely devoid of repetitive DNA elements. Only 1.1% of the assembly consists of mostly non-functional, transposon-derived sequences. This is considerably lower than in most fungi (Supplementary Fig. S2, and Supplementary Tables S3 and S4). Surprisingly, we could not detect any of the known components of the RNA interference pathway, which are thought to have a function in restricting mobile elements¹², and there is no genomic evidence for gene inactivation by repeat-induced point mutation (RIP)¹³. However, it has recently been shown that the expression of heterologous genes in *U. maydis* often results in premature termination of transcription¹⁴. On these grounds, we speculate that *U. maydis* uses this novel mechanism to restrict the activity of invading genes.

Similarly to other fungi and plants^{8,15,16}, centromeres in *U. maydis* seem to coincide with retroelement (HobS)-containing regions that occur once on each chromosome (Supplementary Fig. S1). All known DNA fragments (ARS elements) that allow autonomous replication of plasmids in *U. maydis*¹⁷ match such regions (Supplementary Fig. S1 and Supplementary Table S4). Apparently, the maintenance of plasmids in *U. maydis* requires a centromeric region in addition to an origin of DNA replication, in a similar manner to the situation in *Yarrowia lipolytica*¹⁸. Within the *U. maydis* ARS elements we detected a perfectly conserved 11-bp sequence (ATTCACGATTC) that is strongly over-represented in the genome (5,236 versus 10 expected). Because 96% of these elements are located in intergenic regions, we postulate that this motif defines the origin of replication. In *S. pombe*, functional replication origins occur in comparable numbers and are restricted to intergenic regions. However, these AT-rich regions lack a conserved consensus sequence¹⁹. Even within the basidiomycetes, *U. maydis* is unique in exhibiting such a conserved motif.

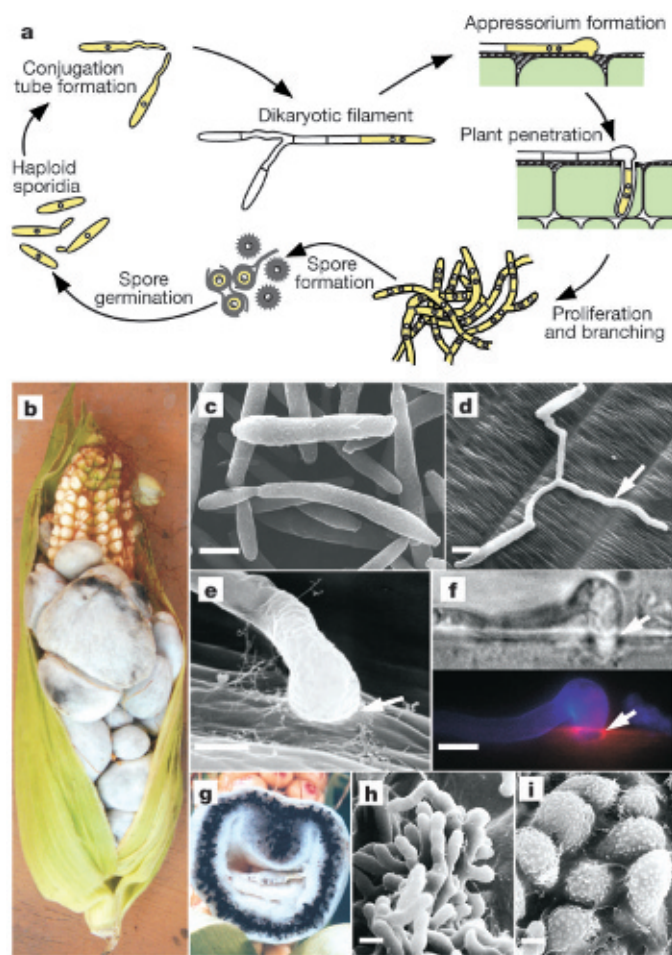


Figure 1 | Life cycle of *U. maydis*. **a**, Developmental stages in the *U. maydis* life cycle. **b**, Tumour formation on maize. **c**, Scanning electron microscopy (SEM) image of haploid sporidia. **d**, SEM image of mated sporidia on plant epidermis; arrow denotes dikaryotic filament. **e**, SEM image of appressorium; arrow marks entry point. **f**, Top, differential interference contrast image of appressorium; bottom, epifluorescence image of fungal cell wall stained with calcofluor (blue) and endocytotic vesicles stained with FM4-64 (red). The bright ring indicates active secretion and endocytosis at the fungus–plant interface; arrows indicate the penetration point. **g**, Black teliospores visible in tumour section. **h**, SEM image of sporogenous hyphae and early stages of spore development. **i**, SEM image of ornamented teliospores. Scale bars, 5 µm.

Remarkably, *U. maydis* possesses only few genes known to be involved in pathogenesis in other fungi. For example, the cereal pathogens *M. grisea*, *F. graminearum* and *Cochliobolus heterostrophus* contain large numbers (15–25) of genes encoding polyketide synthases^{20,21}, enzymes involved in the production of small bioactive compounds such as antibiotics or mycotoxins. In contrast, *U. maydis* contains only three. Genes encoding polysaccharide hydrolases, polysaccharide lyases and pectin esterases are considered to be signatures of necrotrophic fungi that use such enzymes to degrade living and dead plant tissue. *U. maydis* contains only 33 such hydrolytic enzymes, in contrast with 138 and 103 for *M. grisea* and *F. graminearum*, respectively (Supplementary Table S5). The minimal set of hydrolytic enzymes found in *U. maydis* seems perfectly in line with its biotrophic lifestyle, in which damage to the host should be minimized and the release of cell wall fragments, which often trigger plant defence responses, has to be avoided².

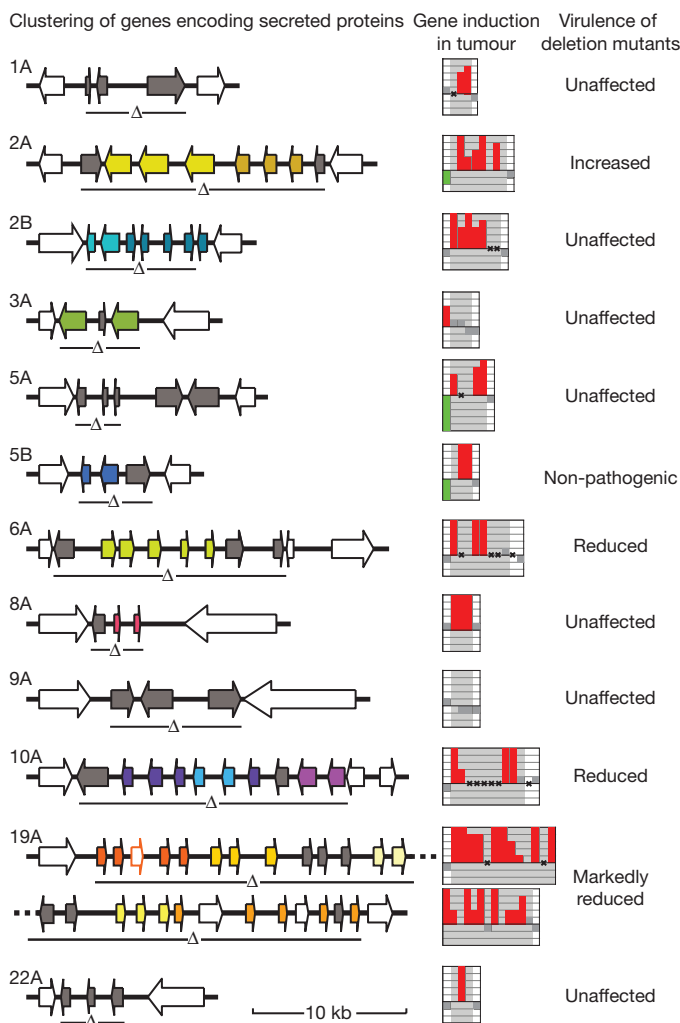


Figure 2 | Gene clusters for secreted proteins. Left, schematic representation of gene clusters and flanking genes. Genes are depicted as arrows, filling reflects the fact that encoded proteins are predicted to be secreted. Colours indicate protein families, unrelated cluster genes are shaded in grey. Regions deleted in cluster mutants are indicated (Δ). Middle, DNA-array analysis of cluster gene expression in tumour tissue. Significant changes in gene expression (compared with axenic culture) were categorized (fold changes on the y-axis: less than 3, 3 to less than 5, 5 to less than 10, 10 to less than 20, at least 20), with red columns indicating induction, green columns repression, and grey columns insignificant changes. Gaps indicate undetectable transcript levels; genes not present on the array are marked with a cross. Light grey areas demarcate clustered genes. See Supplementary Table S6 and Supplementary Fig. S3 for further details.

The paucity of secreted cell-wall-degrading enzymes stands in sharp contrast to the large number of secreted proteins with unknown function. Of 426 proteins predicted to be secreted, 298 (70%) cannot be ascribed a function, and of these almost two-thirds (193) are specific for *U. maydis*. Of all genes encoding secreted proteins, 18.6% are arranged in 12 gene clusters (Fig. 2). The clusters are scattered all over the genome and comprise 3–26 genes. Eight of the 12 clusters contain groups of two to five related genes in tandem arrays, indicating that they might have arisen by duplication. DNA-array analysis revealed that the expression of most clustered genes is induced in tumour tissue, whereas that of flanking genes is not (Fig. 2, and Supplementary Table S6). Although seven of the clustered genes that were induced in tumour tissue were also upregulated by the central regulator of pathogenic development, the bE/bW heterodimer, they were not induced by pheromone stimulation, cyclic AMP signalling, changing of nitrogen or carbon sources, iron depletion, or oxidative stress (data not shown). The specific upregulation of many cluster genes in tumour tissue indicates a possible concerted function during pathogenic development.

To test this assertion, we constructed deletion mutants for all 12 clusters in strain SG200. Mutants were not affected in their growth on minimal medium, showed no morphological alterations and were indistinguishable from strain SG200 in their ability to produce filaments (not shown). Mutants were syringe-injected into maize seedlings. In five cases, deletions caused clear disease-associated phenotypes (Fig. 2, Supplementary Table S6 and Supplementary Fig. S3). Linkage of the observed phenotype to the respective deletion was directly confirmed for cluster 5B by complementation (Supplementary Fig. S4). For the other four clusters, which were significantly larger, complementation attempts were not successful for technical reasons. We therefore generated three independent mutants in each case, which all displayed the same virulence phenotype (Supplementary Fig. S3). Mutants 6A and 10A were still able to infect plants, but the incidence and size of tumours were reduced. Two mutants, 5B and 19A, arrested growth at distinct stages of biotrophic development. Mutant 19A was able to penetrate and to grow inside the plant tissue, but failed to induce large tumours and was defective in spore formation. It is conceivable that some of the proteins encoded by this cluster have tumour-inducing activity. Alternatively, they may suppress plant defence reactions or reprogram the metabolism of the host to allocate resources to the fungal parasite. Deletion of cluster 5B resulted in growth arrest early during penetration of the epidermis, which could indicate a specific need of these proteins during the establishment of a functional interface between the fungus and the host cell. Finally, mutants deleted for cluster 2A showed increased virulence, as judged by the incidence and size of tumours. This hyper-virulence phenotype indicates that the respective proteins might attenuate fungal proliferation. For biotrophic fungi it may be important to prevent the premature development of disease symptoms, because this could affect plant growth so severely that the fungus might not be able to complete its life cycle. Seven cluster mutants were not affected in virulence. For four of these clusters, related genes were found elsewhere in the genome (Supplementary Table S6); the number of clusters with crucial functions for disease progression might therefore actually be even higher.

Our results have shown that secreted protein effectors are essential for fungal proliferation inside the plant host. How these novel effectors exert their function is currently unknown. We envisage that some of the proteins are translocated into plant cells, as has recently been observed in rust and oomycete plant pathogens^{22–24}. These pathogens develop specialized infection structures (haustoria), which are implicated in the exchange of nutrients and proteins². *U. maydis* lacks such structures, but during intracellular growth of the infecting hyphae an extended interaction zone is established, which may be the site at which protein translocation into the host cell takes place.

The genome sequence of the plant pathogenic fungus *U. maydis* has provided unexpected insights into the peculiarities of a biotrophic fungal pathogen. It is apparent that plant cell wall degradation by the fungus is minimized, whereas the secretion of novel protein effectors has a decisive function during infection. This strategy—to live in ‘pretend harmony’ with its host—may be shared not only with other obligate biotrophic pathogens but also with plant-growth-promoting mycorrhizal fungi. The availability of the genome sequence of *U. maydis*, combined with its genetic tractability, therefore offers an excellent opportunity to unravel the molecular secrets of fungal biotrophy. The identification of ‘biotrophy clusters’ in *U. maydis* is likely to have the same inspiring impact on understanding fungal disease strategies as the way in which the discovery of bacterial pathogenicity islands has shaped our present view of bacterial infection strategies.

METHODS

Strains. *Ustilago maydis* 521 (DSMZ number 14603; Supplementary Methods) and FB1 (ref. 7) were used as DNA donors for sequencing. FB2 (ref. 7) was used as the mating partner of FB1 for plant infections. The haploid pathogenic strain SG200 was used for generating deletion mutants and is described in Supplementary Methods.

Genome sequencing and annotation. Strain 521 was sequenced by a combinatorial approach relying on a mapped bacterial artificial chromosome library (2.9× coverage) and a whole-genome shotgun approach (10× coverage). Strain FB1 was sequenced by a whole-genome shotgun approach (5× coverage) (see Supplementary Methods). Predicted protein-encoding genes (see the FGI website) for *U. maydis* were refined manually, including sequence information from more than 4,100 EST clusters (Supplementary Methods) and automatically annotated with the MIPS PEDANT suite²⁵. Data can be accessed on the MUMDB website.

Prediction of secreted proteins. For the prediction of amino-terminal secretion signals, SignalP 3.0 (ref. 26) was used. A total of 750 proteins were predicted to carry a signal peptide both by the hidden Markov and the neural network algorithms. These candidates were analysed with the integral prediction score of ProtComp 6.0 (<http://www.softberry.com>), yielding 426 candidate secreted proteins. In addition, TARGETP²⁷ was used to predict protein localization.

DNA-array analysis. For DNA-array analysis, custom-designed Affymetrix chips were used. Probe sets were designed on the basis of the map-based sequencing assembly. For each predicted gene, 33 perfect match and 33 corresponding mismatch probes were designed, covering a region of 800 bp at the 3′ ends. The *U. maydis* DNA arrays address about 6,200 genes. Probe sets for the individual genes are shown on the MUMDB website. For DNA-array analysis, RNA was extracted from strain FB1 grown to an $A_{600\text{ nm}}$ of 0.5 at 28 °C in liquid array medium (AM), which consisted of 6.25% (v/v) salt solution²⁸, 1% (v/v) vitamin solution²⁸, 30 mM L-glutamine, 1% (w/v) glucose, pH 7.0 (filter sterilized). For RNA from tumour tissue, 7-day-old corn plants were infected with a mixture of strains FB1 and FB2, as described previously²⁹. Tumours were harvested at day 13 after infection. Total RNA was extracted from tumour tissue and axenic cultures with the use of the Trizol method in accordance with the manufacturer’s instructions (Invitrogen). RNA was purified with RNeasy MinElute columns (Qiagen), and RNA integrity was checked on an Agilent Bioanalyser 2100. DNA-array analyses were performed in accordance with the standard Affymetrix protocol in at least two biological replicates. Data analysis was performed with the Affymetrix Micro Array Suite 5.1 software package.

Mutant generation and analysis. Cluster mutants were generated in strain SG200 by gene replacement with PCR-generated constructs as described¹¹, or by subcloning the PCR-derived constructs first. In the latter case, both border fragments were sequenced and shown not to carry mutations. For each cluster, at least two independent mutants were generated and assayed repeatedly for virulence on 7-day-old seedlings of Early Golden Bantam, with a minimum of 40 plants per mutant. Symptom development was scored 12 days after infection. Details of the infection procedure and rating of symptoms are given in Supplementary Fig. S3. Fungal development was monitored by staining with calcofluor and chlorazole black E as described²⁹. Cluster mutant 5B, which is nonpathogenic, was complemented by transformation with a DNA fragment comprising the entire cluster region, ligated to a carboxin resistance cassette. Details are given in Supplementary Fig. S4.

Received 18 August; accepted 15 September 2006.

1. Martinez-Espinoza, A. D., Garcia-Pedrajas, M. D. & Gold, S. E. The Ustilaginales as plant pests and model systems. *Fungal Genet. Biol.* **35**, 1–20 (2002).

2. Mendgen, K. & Hahn, M. Plant infection and the establishment of fungal biotrophy. *Trends Plant Sci.* **7**, 352–356 (2002).
3. Wassenaar, T. M. & Gaastra, W. Bacterial virulence: can we draw the line? *FEMS Microbiol. Lett.* **201**, 1–7 (2001).
4. Bauer, R., Oberwinkler, F. & Vanky, K. Ultrastructural markers and systematics in smut fungi and allied taxa. *Can. J. Bot.* **75**, 1273–1314 (1997).
5. Banuett, F. Genetics of *Ustilago maydis*, a fungal pathogen that induces tumors in maize. *Annu. Rev. Genet.* **29**, 179–208 (1995).
6. Meksem, K. et al. A bacterial artificial chromosome based physical map of the *Ustilago maydis* genome. *Genome* **48**, 207–216 (2005).
7. Banuett, F. & Herskowitz, I. Different *a* alleles of *Ustilago maydis* are necessary for maintenance of filamentous growth but not for meiosis. *Proc. Natl Acad. Sci. USA* **86**, 5878–5882 (1989).
8. Loftus, B. J. et al. The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* **307**, 1321–1324 (2005).
9. Roy, S. W. & Gilbert, W. The evolution of spliceosomal introns: patterns, puzzles and progress. *Nature Rev. Genet.* **7**, 211–221 (2006).
10. Holliday, R. Early studies on recombination and DNA repair in *Ustilago maydis*. *DNA Repair (Amst.)* **3**, 671–682 (2004).
11. Kämper, J. A PCR-based system for highly efficient generation of gene replacement mutants in *Ustilago maydis*. *Mol. Genet. Genomics* **271**, 103–110 (2004).
12. Vastenhouw, N. L. & Plasterk, R. H. RNAi protects the *Caenorhabditis elegans* germline against transposition. *Trends Genet.* **20**, 314–319 (2004).
13. Galagan, J. E. & Selker, E. U. RIP: the evolutionary cost of genome defense. *Trends Genet.* **20**, 417–423 (2004).
14. Zarnack, K. et al. Tetracycline-regulated gene expression in the pathogen *Ustilago maydis*. *Fungal Genet. Biol.* **43**, 727–738 (2006).
15. Cambareri, E. B., Aisner, R. & Carbon, J. Structure of the chromosome VII centromere region in *Neurospora crassa*: degenerate transposons and simple repeats. *Mol. Cell. Biol.* **18**, 5465–5477 (1998).
16. Houben, A. & Schubert, I. DNA and proteins of plant centromeres. *Curr. Opin. Plant Biol.* **6**, 554–560 (2003).
17. Tsukuda, T., Carleton, S., Fotheringham, S. & Holloman, W. K. Isolation and characterization of an autonomously replicating sequence from *Ustilago maydis*. *Mol. Cell. Biol.* **8**, 3703–3709 (1988).
18. Vernis, L. et al. Only centromeres can supply the partition system required for ARS function in the yeast *Yarrowia lipolytica*. *J. Mol. Biol.* **305**, 203–217 (2001).
19. Dai, J., Chuang, R. Y. & Kelly, T. J. DNA replication origins in the *Schizosaccharomyces pombe* genome. *Proc. Natl Acad. Sci. USA* **102**, 337–342 (2005).
20. Dean, R. A. et al. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* **434**, 980–986 (2005).
21. Kroken, S., Glass, N. L., Taylor, J. W., Yoder, O. C. & Turgeon, B. G. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proc. Natl Acad. Sci. USA* **100**, 15670–15675 (2003).
22. Kemen, E. et al. Identification of a protein from rust fungi transferred from haustoria into infected plant cells. *Mol. Plant Microbe Interact.* **18**, 1130–1139 (2005).
23. Dodds, P. N., Lawrence, G. J., Catanzariti, A. M., Ayliffe, M. A. & Ellis, J. G. The *Melampsora lini* AvrL567 avirulence genes are expressed in haustoria and their products are recognized inside plant cells. *Plant Cell* **16**, 755–768 (2004).
24. Birch, P. R., Rehmany, A. P., Pritchard, L., Kamoun, S. & Beynon, J. L. Trafficking arms: oomycete effectors enter host plant cells. *Trends Microbiol.* **14**, 8–11 (2006).
25. Mewes, H. W. et al. MIPS: analysis and annotation of proteins from whole genomes. *Nucleic Acids Res.* **32**, D41–D44 (2004).
26. Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795 (2004).
27. Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005–1016 (2000).
28. Holliday, R. in *Handbook of Genetics* (ed. King, R. C.) 575–595 (Plenum, New York, 1974).
29. Brachmann, A., Schirawski, J., Müller, P. & Kahmann, R. An unusual MAP kinase is required for efficient penetration of the plant surface by *Ustilago maydis*. *EMBO J.* **22**, 2199–2210 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements J.K., M. B. and R.K. thank G. Sawers and U. Kämper for critical reading of the manuscript. The genome sequencing of *Ustilago maydis* strain 521 is part of the fungal genome initiative and was funded by National Human Genome Research Institute (USA) and BayerCropScience AG (Germany). F.B. was supported by a grant from the National Institutes of Health (USA). J.K. and R.K. thank the German Ministry of Education and Science (BMBF) for financing the DNA array setup and the Max Planck Society for their support of the manual genome annotation. F.B. was supported by a grant from the National Institutes of Health, B.J.S. was supported by the Natural Sciences and Engineering Research Council of Canada and the Canada Foundation for Innovation, J.W.K. received funding from the Natural Sciences and Engineering Research Council of Canada,

J.R.-H. received funding from CONACYT, México, A.M.-M. was supported by a fellowship from the Humboldt Foundation, and L.M. was supported by an EU grant.

Author Contributions All authors were involved in planning and executing the genome sequencing project. B.W.B., J.G., L.-J.M., E.W.M., D.D., C.M.W., J.B., S.Y., D.B.J., S.C., C.N., E.K., G.F., P.H.S., I.H.-H., M. Vaupel, H.V., T.S., J.M., D.P., C.S., A.G., F.C. and V. Vysotskaia contributed to the three independent sequencing projects; M.M., G.M., U.G., D.H., M.O. and H.-W.M. were responsible for gene model refinement, database design and database maintenance; G.M., J. Kämper, R.K., G.S., M. Feldbrügge, J.S., C.W.B., U.F., M.B., B.J.S., M.J.C., E.C.H.H., S.M., F.B., J.W.K., K.J.B., J. Klose, S.E.G., S.J.K., M.H.P., H.A.B.W., R.deV., H.J.D., J.R.-H., C.G.R.-P., L.O.-C., M.McC., K.S., J.P.-M., J.I.I., W.H., P.G., P.S.-A., M. Farman, J.E.S., R.S., J.M.G.-P., J.C.K., W.L. and D.H. were involved in functional annotation and

interpretation; T.B., O.M., L.M., A.M.-M., D.G., K.M., N.R., V. Vincon, M. Vraneš, M.S. and O.L. performed experiments. J. Kämper, R.K. and M.B. wrote and edited the paper with input from L.-J.M., J.G., F.B., J.W.K., B.J.S. and S.E.G. Individual contributions of authors can be found as Supplementary Notes.

Author Information The whole-genome shotgun data have been deposited at GenBank under the project accession number AACP00000000. Individual, automatically generated gene models have been deposited under accession numbers XM_751055 to XM_757567. Supplementary Table S8 lists all gene models that have been altered during manual reannotation. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.K. (kahmann@mpi-marburg.mpg.de), M.B. (boelker@staff.uni-marburg.de) or J.K. (kaemper@mpi-marburg.mpg.de).

LETTERS

Two modes of fusion pore opening revealed by cell-attached recordings at a synapse

Liming He¹, Xin-Sheng Wu¹, Raja Mohan¹ & Ling-Gang Wu¹

Fusion of a vesicle with the cell membrane opens a pore that releases transmitter to the extracellular space^{1–3}. The pore can either dilate fully so that the vesicle collapses completely, or close rapidly to generate ‘kiss-and-run’ fusion^{1,2,4–7}. The size of the pore determines the release rate². At synapses, the size of the fusion pore is unclear, ‘kiss-and-run’ remains controversial^{8–15}, and the ability of ‘kiss-and-run’ fusion to generate rapid synaptic currents^{16,17} is questionable¹⁸. Here, by recording fusion pore kinetics during single vesicle fusion, we found both full collapse and ‘kiss-and-run’ fusion at calyx-type synapses. For full collapse, the initial fusion pore conductance (G_p) was usually >375 pS and increased rapidly at ≥ 299 pS ms⁻¹. ‘Kiss-and-run’ fusion was seen as a brief capacitance flicker (<2 s) with $G_p > 288$ pS for most flickers, but within 15–288 pS for the remaining flickers. Large G_p (>288 pS) might discharge transmitter rapidly and thereby cause rapid synaptic currents, whereas small G_p might generate slow and small synaptic currents. These results show that ‘kiss-and-run’ fusion occurs at synapses and that it can generate rapid postsynaptic currents, and suggest that various fusion pore sizes help to control the kinetics and amplitude of synaptic currents.

We developed a technique to expose release sites in a large nerve terminal, the calyx of Held, which contains clear-core glutamatergic vesicles (Fig. 1a). The calyx membrane in apposition to the postsynaptic neuron, where release presumably occurs, was exposed by sucking and pulling away the postsynaptic neuron using an electrode (Fig. 1a). This procedure did not significantly change the morphology of the calyx when it was labelled with Lucifer yellow ($n = 8$, Fig. 1a). At the exposed presynaptic membrane, called position I, cell-attached recordings of the imaginary (Im, capacitance) and the real (Re, conductance) components of the complex admittance were made with an electrode (containing 2 mM Ca²⁺, Fig. 1a).

The application of a concentrated potassium solution (including 25 mM KCl and 2 mM CaCl₂) induced, within 10–30 s, discrete capacitance up-steps (for example, Fig. 1b, c) at 0.088 ± 0.02 Hz (Fig. 2a) in 43 out of 124 patches, which was significantly higher than that at rest (0.008 ± 0.002 Hz, $n = 43$ patches, t -test, $p < 0.01$, Figs 1b, 2a; bath solution contained 2.5 mM KCl, 2 mM CaCl₂). When exocytosis was blocked by removing Ca²⁺ from the bath and pipette, high potassium application failed to increase the up-step frequency ($n = 21$ patches, Fig. 1d; see Supplementary Information II). Similarly, when exocytosis was abolished by the protease trypsin, the high-potassium solution failed to increase the up-step frequency ($n = 23$ patches, see Supplementary Information III). This indicates that up-steps were not caused by a capacitance artefact observed in the whole-cell configuration¹², because trypsin blocked exocytosis but not the artefact (Supplementary Information III). Up-steps were not seen during high potassium application at positions not apposing the postsynaptic neuron, called position II (Fig. 1e), which presumably does not contain release sites ($n = 30$ patches; for example,

Fig. 1e). Thus, capacitance up-steps paralleled exocytosis in location and sensitivity to Ca²⁺ and trypsin, indicating that they are caused by exocytosis.

As the calyx could be visualized after removing the postsynaptic neuron (Fig. 1a), the patch at position I was at either the calyx or the residual postsynaptic membrane. Up-steps were recorded from calyces for the following three reasons. First, up-steps were not observed at postsynaptic neurons ($n = 20$ patches). Second, after 5–20-min recordings in 43 cell-attached patches that showed up-steps, 32 patches switched, automatically or by suction, to the whole-cell mode at the calyx (11 patches lost seal; see Supplementary Information IV). Third, in an additional 15 cell-attached patches at position I that were recorded using an intracellular pipette solution (see whole-cell recording solutions in Supplementary Information VIII), suction led to the whole-cell configuration in 10 patches. In these 10 whole-cell recordings, a 20-ms depolarization (–80 to +10 mV) evoked a capacitance jump (198 ± 6 fF, $n = 10$; for example, Fig. 1f, black) similar to that (238 ± 17 fF, $n = 10$; for example, Fig. 1f, grey) evoked through a whole-cell electrode at position II. Thus, at least 67% (10/15) of patches at position I were at the calyx, and removal of the postsynaptic neuron did not affect release.

Three pieces of evidence indicate that up-steps reflected the fusion of single vesicles. First, the up-step frequency was ~ 0.088 Hz (Fig. 2a), which made independent, multiple vesicle fusion highly unlikely. Second, the up-step frequency increase induced by high potassium (~ 11 times, Fig. 2a) fell within the range (6–32 times, mean 18 ± 5.8 times, $n = 4$; Fig. 2b) of the increase in the frequency of miniature excitatory postsynaptic currents (mEPSCs) that was induced by high potassium (25 mM). Third, the size distribution of up-steps was similar to that predicted from electron microscopy. Given a specific membrane capacitance of 8–11 fF μm^{-2} (ref. 21), a vesicle diameter of 30–80 nm, as found in calyces^{19,20}, predicts a membrane capacitance of 23–221 aF for one vesicle. Similarly, the amplitude of most up-steps was 30–220 aF (Fig. 2c). The mean vesicle diameter was 46 nm (ref. 20), predicting a mean vesicle capacitance of 53–73 aF. Similarly, the mean of up-steps with an amplitude <220 aF was 73.3 ± 1.3 aF ($n = 562$ up-steps). A small fraction of up-steps were >220 aF (Fig. 2c). As large, dense-core vesicles are rare in calyces^{19,20}, these large up-steps might be caused by independent or coordinated multi-vesicular fusion²². Consistent with this possibility, the amplitude of all ‘kiss-and-run’ (see definition later) up-steps, which reflect single vesicle fusion, was <220 aF (Fig. 2f, g) with a mean of 74.0 ± 3.2 aF ($n = 114$). Our further analysis therefore did not include up-steps >220 aF.

Similar to up-steps, most down-steps were <220 aF (Fig. 2d, $n = 215$ down-steps). The amplitude distribution of down-steps that were not immediately preceded by an up-step (that is, not part of a kiss-and-run sequence; see below) was broader (Fig. 2d) than that of up-steps (Fig. 2c; see also ref. 18). These down-steps might reflect the

¹National Institute of Neurological Disorders and Stroke, 35 Convent Drive, Building 35, Room 2B-1012, Bethesda, Maryland 20892, USA.

retrieval of vesicles that are similar but not identical to vesicles undergoing exocytosis¹⁸.

Out of 562 up-steps, 116 were followed by a down-step within 2 s. Except for two down-steps, the amplitude of down-steps was <20% different from their preceding up-steps (Fig. 2f). The relationship between down- and up-step amplitude was linear (Fig. 2f, slope = 0.99). By contrast, the amplitude of down-steps was not correlated with that of an unpaired up-step occurring earlier or later (slope = 0.12, not shown; see also Supplementary Information V). These results indicate that a pair of successive up- and down-steps with similar amplitudes, called a capacitance flicker^{2,7}, arose from the same vesicle and thus reflected 'kiss-and-run' exocytosis⁶. The flicker duration ranged from 10 to 1,953 ms with a mean of 273 ± 22 ms ($n = 114$, Fig. 2h).

Next, we measured the fusion pore conductance (G_p) based on

$$G_p = (\Delta Re^2 + \Delta Im^2) / \Delta Re \quad (1)$$

where Δ means the difference from baseline². G_p can be resolved when ΔRe is detectable. In 17% (19/114) of flickers, ΔRe was detected, and G_p was 66 ± 7 pS (range: 15–110 pS, $n = 19$ flickers;

for example, Fig. 3a). The detected ΔRe was not due to phase adjustment errors because non-flicker events were not accompanied by a persistent ΔRe (Fig. 3a; see Supplementary Information VIII). In 83% (95/114) of flickers, ΔRe was not detected (Fig. 3b). When these flickers were averaged, the mean ΔRe was also not detectable (Fig. 3c), indicating that most flickers had a mean $G_p > 288$ pS (see Methods for calculation).

Up to this point, G_p was measured at the sine wave frequency of 20 kHz. To resolve larger G_p , we increased the sine wave frequency to 90 kHz (ref. 23). In this condition, G_p could be measured in 15% (5/32) of flickers. The measurable G_p was 146 ± 24 pS ($n = 5$ flickers, range: 103–237 pS; for example, Fig. 3d), which was larger than the detectable mean G_p (66 ± 7 pS, $n = 19$) obtained at 20 kHz of sine wave. This result was expected because increasing the sine wave frequency increased both the upper and the lower G_p detection limit²³. In summary, G_p during flickers ranged from 15 pS to >288 pS, but with most >288 pS.

For non-flicker up-steps measured at 20 kHz of sine wave, ~98.6% (442 out of 448 up-steps) were not accompanied by a detectable transient ΔRe (for example, Figs 1b, c, 3a; see Supplementary Information VI for the remaining 1.4%). Even after averaging all up-steps ($n = 448$), a transient ΔRe was not detectable (Fig. 3e), indicating that the initial G_p was >375 pS and/or too fast to resolve (see Methods for calculation). At 90 kHz of sine wave, G_p was resolved in 10% (10/102) of non-flicker up-steps (Fig. 3f). The initial G_p was 256 ± 32 pS ($n = 10$), which increased at a rate of $0\text{--}89$ pS ms⁻¹ for ~10–320 ms, followed by a rapid increase of G_p at 299 ± 35 pS ms⁻¹. These values were the lower limit for most non-flicker up-steps in which G_p was too large to resolve.

We have shown that ~80% (448/562) of up-steps were non-flickers or full-collapse exocytosis with a mean initial $G_p > 375$ pS, whereas ~20% of up-steps were flickers, among which ~83% had

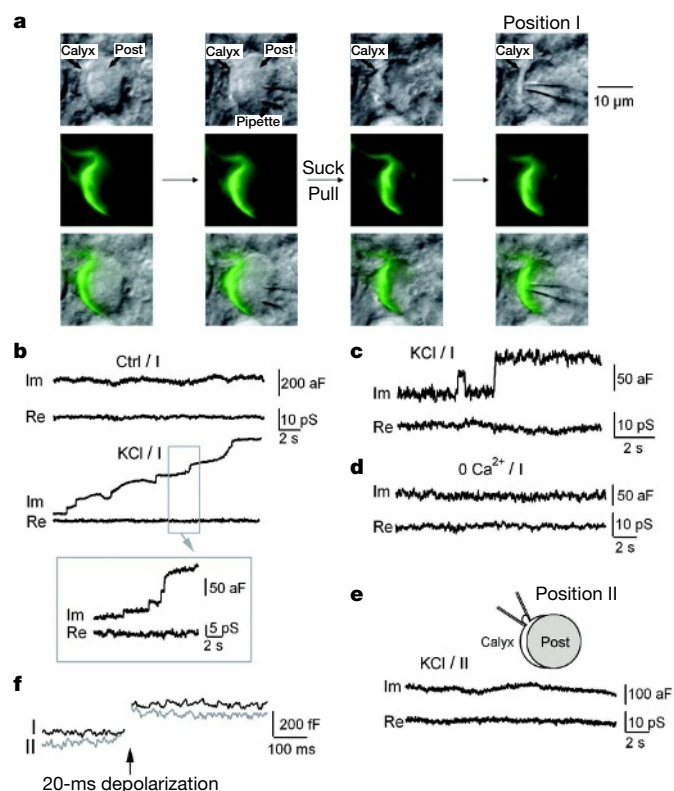


Figure 1 | Capacitance up-steps at the nerve terminal. **a**, Left: DIC (differential infrared contrast, upper), fluorescence (middle), and superimposed (lower) images of a calyx associated with a postsynaptic neuron (post). The fluorescence was obtained by dialysing Lucifer yellow (4 mg ml^{-1} , 3 min) into the calyx through a whole-cell pipette. The pipette was withdrawn. Middle left and right: the postsynaptic neuron was sucked and pulled away by a pipette shown in the middle left. Right: cell-attached recording at the release face (position I) of the calyx membrane. **b**, Im and Re traces from cell-attached recordings at position I in control conditions (Ctrl, upper) and during application of 25 mM KCl (lower; scale bars in upper panel). The inset shows discrete up-steps. **c**, More Im and Re traces at position I during application of 25 mM KCl. **d**, Traces obtained in similar conditions as in panels **b** and **c**, but without Ca^{2+} in the bath and the pipette solution. **e**, Im and Re traces during high potassium application from position II of the calyx (see drawing). **f**, Capacitance jumps induced by a 20-ms depolarization from -80 to $+10$ mV (arrow) at the whole-cell configuration with electrodes placed at position I (black) and II (grey).

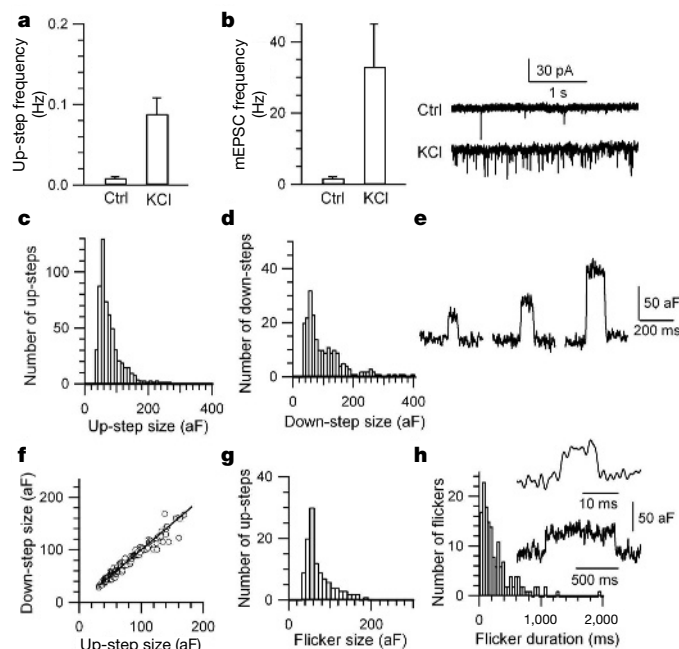


Figure 2 | Capacitance up-steps reflect single vesicle fusion. **a**, **b**, The up-step (**a**) and the mEPSC (**b**) frequency in control conditions (Ctrl) and during high potassium (25 mM) application. Sampled mEPSCs before (Ctrl) and during high potassium application are also shown (**b**). Error bars show s.e.m. **c**, Amplitude distribution of up-steps, including flickers and non-flickers (from 43 patches). **d**, Amplitude distribution of down-steps not preceded by an up-step within 2 s (from 43 patches). **e**, Sampled capacitance flickers. **f**, The relationship between down- and up-step size of flickers was fit with a regression line (slope = 0.99). **g**, Flicker up-step size distribution. **h**, Flicker duration distribution. Two samples are also shown.

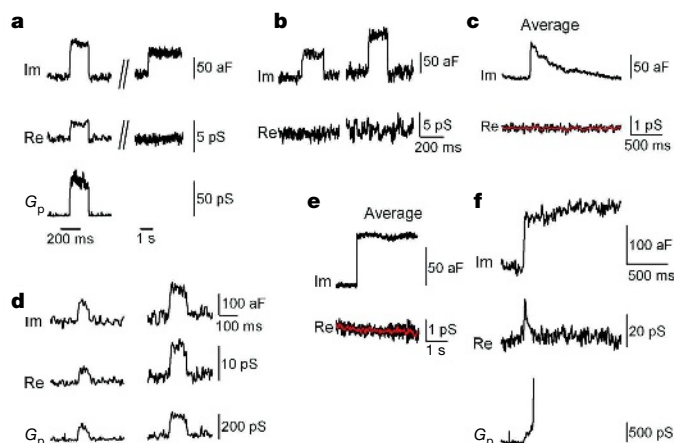


Figure 3 | Modes of fusion pore openings. **a**, I_m , R_e and G_p during a capacitance flicker. A non-flicker up-step (right) occurring 10 s later was not accompanied by detectable R_e changes, indicating proper phase adjustment. **b**, Sampled I_m and R_e of two flickers without detectable R_e changes. **c**, Averaged I_m and R_e from 95 flickers not accompanied by detectable R_e (black). The R_e trace was also low-pass filtered at 10 Hz (red) to reduce the noise. **d**, I_m , R_e and G_p during two flickers recorded at 90 kHz of sine wave voltage command (if not mentioned, the frequency was 20 kHz). **e**, I_m and R_e traces averaged from 448 non-flicker up-steps. **f**, I_m , R_e and G_p during a non-flicker up-step recorded at 90 kHz of sine wave.

an initial $G_p > 288$ pS. Thus, $\sim 97\%$ ($80\% + 20\% \times 83\%$) of fusion events had an initial $G_p > 288$ pS, whereas $\sim 3\%$ ($20\% \times 17\%$) were flickers with a mean G_p of 66 pS. A $G_p > 288$ pS in 97% of up-steps might correspond to a fusion pore diameter (D_p) > 2.3 nm (equation (2) in Methods), which could allow discharge of glutamate with a time constant (τ_{glu_v}) < 0.54 ms (equation (3) in Methods). ‘Kiss-and-run’ fusion with a mean G_p of 66 pS in $\sim 3\%$ of fusion events corresponds to a D_p of 1.1 nm and a τ_{glu_v} of 2.3 ms.

Simulation using the MCell 2.50 program (Supplementary Information VII)²⁴ predicted that vesicle fusion with a G_p of 288 pS– ∞ , as seen in $\sim 97\%$ of up-steps, would cause an mEPSC with a 10–90% rise time of 0.15–0.36 ms (Supplementary Fig. 8a, Supplementary Table 1). By contrast, ‘kiss-and-run’ fusion with a G_p of ~ 66 pS, as seen in $\sim 3\%$ of fusion events, would cause an mEPSC with a much slower rise and decay and a smaller amplitude (Supplementary Fig. 8d, Supplementary Table 1). Consistent with these predictions, the mean of all experimentally measured mEPSCs had a rapid 10–90% rise time (0.21–0.27 ms) within the predicted range (Supplementary Fig. 8b, c, Supplementary Information VII). A very small fraction of measured mEPSCs were small in amplitude and slow in both rise and decay (Supplementary Fig. 8e, Supplementary Information VII). These results indicate that the initial fusion pore size might help to control the rise, decay and peak amplitude of the mEPSC.

This work studied fusion modes at synapses. ‘Kiss-and-run’ fusion provides a mechanism for rapid recycling of vesicles²⁵. However, whether it exists at synapses is intensely debated, partly because fusion pore size is difficult to resolve at synapses (Supplementary Information I)^{8–15}. Here, using cell-attached recordings at the release face of calyceal terminals, we found two fusion modes, full collapse and ‘kiss-and-run’. ‘Kiss-and-run’ fusion was identified as capacitance flickers caused by single vesicle fusion and retrieval at a small patch of membrane. The flicker duration was usually 10 ms–1 s (Fig. 2h), and the pore diameter was mainly > 2.3 nm. These results provide strong evidence for ‘kiss-and-run’ fusion at synapses.

‘Kiss-and-run’ fusion allows transmitter to be released from large, dense-core vesicles^{2,3}. Would it release transmitter from small synaptic vesicles? A study of micro-vesicles, which resemble but are not synaptic vesicles, showed a G_p (19 pS) that was too small to release transmitter rapidly¹⁸, casting doubt on the ability of ‘kiss-and-run’

fusion to generate the sub-millisecond rise in postsynaptic currents that can be seen at synapses^{16,17} (Supplementary Fig. 8b, c). Unlike this study of micro-vesicles, we found that G_p during ‘kiss-and-run’ fusion at calyx-type synapses was mostly > 288 pS, which would allow sub-millisecond transmitter discharge and generation of a rapid and large synaptic current (Supplementary Fig. 8a–c, Supplementary Table 1). We conclude that ‘kiss-and-run’ fusion can mediate fast synaptic release.

Although full-collapse fusion is well established^{2,4}, its fusion pore kinetics are unclear at synapses. We found that the initial G_p of full-collapse fusion was usually > 375 pS at calyx-type synapses, which allows rapid transmitter discharge and generation of fast synaptic currents (Supplementary Figs 7a, 8a). Previous studies proposed that the size and kinetics of the fusion pore control quantal size and amplitude at synapses^{18,26,27}. Our measurements of the fusion pore conductance provide direct support for this proposal. We found that most fusion events had a large G_p , which could generate rapid and large synaptic currents and thus explain why most mEPSCs at calyx-type synapses rise rapidly and why fast synaptic currents are widely observed^{16,17} (Supplementary Fig. 8b, c). Some ‘kiss-and-run’ fusion events had a small G_p , which can account for slower and smaller mEPSCs at synapses^{18,26,27} (Supplementary Fig. 8e). Thus, fusion pore size and fusion modes might help to control the amplitude and the kinetics of postsynaptic currents.

METHODS

Parasagittal brainstem slices (200 μ m thick) containing the medial nucleus of the trapezoid body were prepared from 6–8-day-old Wistar rats with a vibratome¹¹. Cell-attached capacitance measurements were performed with an SR830 2-phase lock-in amplifier (Stanford Research Systems) coupled to an EPC-8 patch-clamp amplifier (HEKA Electronics)^{2,18}. A sine wave with a root-mean-square (r.m.s.) amplitude of 200 mV at 20 kHz (if not specified) or 90 kHz was superimposed on a command potential of 0 mV. Bath and pipette solutions, detailed methods of cell-attached capacitance measurements, phase adjustment, and detection and measurement of capacitance up- and down-steps are described in Supplementary Information VIII. Results are presented as means $\pm$ standard errors. In all figures, I_m and R_e traces are low-pass filtered at 50 Hz.

We did not detect ΔR_e in the mean R_e trace (Fig. 3c) averaged from 95 flickers, each of which showed no detectable ΔR_e . This result indicates that G_p was too large to resolve. To estimate the upper detection limit of G_p , we low-pass filtered the mean R_e trace at 10 Hz (Fig. 3c, red). The R_e noise (r.m.s.) reached 0.10 pS (Fig. 3c, red). The detection limit of ΔR_e , set as three times R_e noise, was 0.30 pS. According to equation (1), this value, together with the averaged ΔI_m of 74 aF (Fig. 3c), yielded an upper detection limit for G_p of 288 pS. Thus, the mean G_p for most flickers was > 288 pS. Similar methods were used to calculate the upper detection limit of G_p for non-flicker up-steps (Fig. 3e). With the G_p value calculated from equation (1), we estimated the pore diameter (D_p) and the time constant (τ_{glu_v}) of discharging glutamate out of the vesicle by diffusion based on^{28,29}:

$$D_p = [4G_p\rho\lambda/\pi]^{0.5} \quad (2)$$

$$\tau_{glu_v} = (\pi D_v^3/6)/(\rho K_d G_p) \quad (3)$$

where ρ is the saline resistivity (100 Ω cm); λ , the pore length, is taken as the length of a gap junction channel (15 nm); D_v , the vesicle diameter, is, on average, 46 nm in calyces²⁰; and K_d , the diffusion constant, is $\sim 3.3 \times 10^{-6}$ cm² s⁻¹ for glutamate in the synaptic cleft³⁰.

Received 11 May; accepted 14 September 2006.

Published online 25 October 2006.

1. Breckenridge, L. J. & Almers, W. Currents through the fusion pore that forms during exocytosis of a secretory vesicle. *Nature* **328**, 814–817 (1987).
2. Lindau, M. & Alvarez de Toledo, G. The fusion pore. *Biochim. Biophys. Acta* **164**, 167–173 (2003).
3. Alvarez, D. T., Fernandez-Chacon, R. & Fernandez, J. M. Release of secretory products during transient vesicle fusion. *Nature* **363**, 554–558 (1993).
4. Heuser, J. E. & Reese, T. S. Evidence for recycling of synaptic vesicle membrane during transmitter release at the frog neuromuscular junction. *J. Cell Biol.* **57**, 315–344 (1973).

5. Ceccarelli, B., Hurlbut, W. P. & Mauro, A. Turnover of transmitter and synaptic vesicles at the frog neuromuscular junction. *J. Cell Biol.* **57**, 499–524 (1973).
6. Fesce, R., Grohovaz, F., Valtorta, F. & Meldolesi, J. Neurotransmitter release, fusion or 'kiss and run'? *Trends Cell Biol.* **4**, 1–4 (1994).
7. Fernandez, J. M., Neher, E. & Gomperts, B. D. Capacitance measurements reveal stepwise fusion events in degranulating mast cells. *Nature* **312**, 453–455 (1984).
8. Aravanis, A. M., Pyle, J. L. & Tsien, R. W. Single synaptic vesicles fusing transiently and successively without loss of identity. *Nature* **423**, 643–647 (2003).
9. Gandhi, S. P. & Stevens, C. F. Three modes of synaptic vesicular recycling revealed by single-vesicle imaging. *Nature* **423**, 607–613 (2003).
10. Fernandez-Alfonso, T. & Ryan, T. A. The kinetics of synaptic vesicle pool depletion at CNS synaptic terminals. *Neuron* **41**, 943–953 (2004).
11. Sun, J. Y., Wu, X. S. & Wu, L. G. Single and multiple vesicle fusion induce different rates of endocytosis at a central synapse. *Nature* **417**, 555–559 (2002).
12. Yamashita, T., Hige, T. & Takahashi, T. Vesicle endocytosis requires dynamin-dependent GTP hydrolysis at a fast CNS synapse. *Science* **307**, 124–127 (2005).
13. Harata, N. C., Choi, S., Pyle, J. L., Aravanis, A. M. & Tsien, R. W. Frequency-dependent kinetics and prevalence of kiss-and-run and reuse at hippocampal synapses studied with novel quenching methods. *Neuron* **49**, 243–256 (2006).
14. Wienisch, M. & Klingauf, J. Vesicular proteins exocytosed and subsequently retrieved by compensatory endocytosis are nonidentical. *Nature Neurosci.* **9**, 1019–1027 (2006).
15. Fernandez-Alfonso, T., Kwan, R. & Ryan, T. A. Synaptic vesicles interchange their membrane proteins with a large surface reservoir during recycling. *Neuron* **51**, 179–186 (2006).
16. Stiles, J. R., Van Helden, D., Bartol, T. M. Jr, Salpeter, E. E. & Salpeter, M. M. Miniature endplate current rise times less than 100 microseconds from improved dual recordings can be modeled with passive acetylcholine diffusion from a synaptic vesicle. *Proc. Natl Acad. Sci. USA* **93**, 5747–5752 (1996).
17. Bruns, D. & Jahn, R. Real-time measurement of transmitter release from single synaptic vesicles. *Nature* **377**, 62–65 (1995).
18. Klyachko, V. A. & Jackson, M. B. Capacitance steps and fusion pores of small and large-dense-core vesicles in nerve terminals. *Nature* **418**, 89–92 (2002).
19. Lenn, N. J. & Reese, T. S. The fine structure of nerve endings in the nucleus of the trapezoid body and the ventral cochlear nucleus. *Am. J. Anat.* **118**, 375–390 (1966).
20. Sätzler, K. *et al.* Three-dimensional reconstruction of a calyx of Held and its postsynaptic principal neuron in the medial nucleus of the trapezoid body. *J. Neurosci.* **22**, 10567–10579 (2002).
21. Gentet, L. J., Stuart, G. J. & Clements, J. D. Direct measurement of specific membrane capacitance in neurons. *Biophys. J.* **79**, 314–320 (2000).
22. Singer, J. H., Lassoova, L., Vardi, N. & Diamond, J. S. Coordinated multivesicular release at a mammalian ribbon synapse. *Nature Neurosci.* **7**, 826–833 (2004).
23. Debus, K. & Lindau, M. Resolution of patch capacitance recordings and of fusion pore conductances in small vesicles. *Biophys. J.* **78**, 2983–2997 (2000).
24. Stiles, J. R., Bartol, T. B., Salpeter, M. M., Salpeter, E. E. & Sejnowski, T. J. in *Synapses* (eds Cowan, T. C., Sudhof, T. C. & Stevens, C. F.) 681–732 (Johns Hopkins Univ. Press, Baltimore, 2000).
25. Harata, N. *et al.* Limited numbers of recycling vesicles in small CNS nerve terminals: implications for neural signaling and vesicular cycling. *Trends Neurosci.* **24**, 637–643 (2001).
26. Choi, S., Klingauf, J. & Tsien, R. W. Fusion pore modulation as a presynaptic mechanism contributing to expression of long-term potentiation. *Phil. Trans. R. Soc. Lond. B* **358**, 695–705 (2003).
27. Pawlu, C., DiAntonio, A. & Heckmann, M. Postfusional control of quantal current shape. *Neuron* **42**, 607–618 (2004).
28. Almers, W. *et al.* Millisecond studies of single membrane fusion events. *Ann. NY Acad. Sci.* **635**, 318–327 (1991).
29. Spruce, A. E., Breckenridge, L. J., Lee, A. K. & Almers, W. Properties of the fusion pore that forms during exocytosis of a mast cell secretory vesicle. *Neuron* **4**, 643–654 (1990).
30. Nielsen, T. A., Digregorio, D. A. & Silver, R. A. Modulation of glutamate mobility reveals the mechanism underlying slow-rising AMPAR EPSCs and the diffusion coefficient in the synaptic cleft. *Neuron* **42**, 757–771 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank V. Klyachko and M. Lindau for technical guidance on cell-attached recordings. We thank M. Lindau for the software for data analysis. We thank J. Diamond, W. Wu, L. Xue, W. Grimes, M. Diaz-Bustamante and B. McNeil for help with simulation and data analysis. We thank J. Diamond, D. Nees, K. Paradiso and J. Xu for comments on the manuscript. This work was supported by the National Institute of Neurological Disorders and Stroke Intramural Research Program.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.G.W. (wul@ninds.nih.gov).

LETTERS

Movement of 'gating charge' is coupled to ligand binding in a G-protein-coupled receptor

Yair Ben-Chaim^{1†}, Baron Chanda², Nathan Dascal³, Francisco Bezanilla⁴, Itzhak Parnas¹ & Hanna Parnas¹

Activation by agonist binding of G-protein-coupled receptors (GPCRs) controls most signal transduction processes¹. Although these receptors span the cell membrane, they are not considered to be voltage sensitive. Recently it was shown that both the activity of GPCRs^{2–5} and their affinity towards agonists⁶ are regulated by membrane potential. However, it remains unclear whether GPCRs intrinsically respond to changes in membrane potential. Here we show that two prototypical GPCRs, the m2 and m1 muscarinic receptors (m2R and m1R), display charge-movement-associated currents analogous to 'gating currents' of voltage-gated channels. The gating charge–voltage relationship of m2R correlates well with the voltage dependence of the affinity of the receptor for acetylcholine. The loop that couples m2R and m1R to their G protein has a crucial function in coupling voltage sensing to agonist-binding affinity. Our data strongly indicate that GPCRs serve as sensors for both transmembrane potential and external chemical signals.

Using heterologous expression in *Xenopus* oocytes, we have recently discovered that the apparent binding affinity of acetylcholine (ACh) to two GPCRs, m2R and m1R, is voltage dependent⁶. This was demonstrated by measuring m1R-activated and m2R-activated ion-channel currents and ACh binding to these receptors. Remarkably, depolarization decreases the apparent ACh affinity of m2R but increases that of m1R; the voltage sensitivity does not reside downstream of receptor activation by ACh⁶.

How can GPCRs respond to membrane potential (V_m)? Movement of charge is a fundamental mechanism by which proteins couple changes in membrane electric field to a conformational change. 'Gating current' is due to charge movement driven by changes in V_m in such a protein. Originally predicted by Hodgkin and Huxley⁷, gating currents have been detected in several voltage-gated ion channels⁸, in ion transporters^{9,10} and in a voltage-dependent phosphatase¹¹, but never in a GPCR. We posit that muscarinic GPCRs must also move 'gating' charge in response to changes in V_m . We measured gating currents of these receptors in *Xenopus* oocytes by using a cut-open oocyte voltage-clamp setup¹². The voltage pulse protocol is shown in Fig. 1a. Symmetric capacitive currents, associated with charging of the bilayer, were eliminated by the divided-pulse procedure¹³. Oocytes expressing m2R showed large transient currents in response to depolarizing steps from a prepulse of -120 mV (Fig. 1a) that were not present in water-injected oocytes (Fig. 1b).

At small depolarizations, such as -90 mV, the ON gating current is characterized by two components, a fast one followed by a slow one (arrowhead), which has both a rising and a decaying phase (Fig. 1c). The rising phase of the slow ON gating currents disappears at larger depolarizations (for example $+10$ mV). The kinetics of the fast component (in the range of a few hundreds of microseconds; Fig. 1c inset) suggests that even a single action potential would move

the fast gating charge associated with m2R. To move the slow charge, a train of action potentials may be needed. The decay of the ON gating current was fitted with a two-exponential function. The voltage dependence of the time constant of the slow component was bell-shaped, indicating a two-state process, whereas the fast component showed very little voltage dependence (Supplementary Fig. 1a). The OFF gating currents (Fig. 1d), elicited on returning from a test pulse, show a prominent slow rising phase at all potentials (arrowhead).

The observed ON and OFF currents were not ionic, because they were measured in internal and external solutions composed of non-permeable ions and were not affected by tetraethylammonium. In addition, the integrals of the ON and OFF gating currents (Supplementary Fig. 1b) were similar at all voltages and showed an excellent linear correlation (Supplementary Fig. 1c), as expected for currents associated with charge movement. The dependence of the ON gating charge on V_m (the $Q-V$ curve) showed a characteristic sigmoid shape and was fitted to a two-state model (Boltzmann equation, equation (4) in Supplementary Methods) with the parameters V_{50} (voltage at half maximal gating charge) = -44 mV and z (amount of charge that moves per molecule) = 0.85 (Fig. 2b, filled triangles).

To determine whether the gating currents are linked with the voltage-dependent affinity of m2R, we measured dose-response curves at various values for V_m between -100 and $+40$ mV, using the G-protein-activated K^+ channel (GIRK) as a reporter⁶ (Fig. 2a). m2R shows two affinity states^{6,14,15}, defined by fractional occupancies R^H and R^L . Depolarization causes a transition from the high-affinity to the low-affinity state^{6,16}. This process can be described by the two-state Boltzmann equation. As seen in Fig. 2b, the dependence of R^L on V_m (R^L-V curve; open circles) correlated well with the $Q-V$ relation (filled triangles) and showed similar parameters of fit to the Boltzmann equation, strongly suggesting that the two processes are tightly coupled.

We expected that m1R, which also shows voltage-dependent ACh binding, although opposite to that of m2R⁶, should have gating currents. The expression of m1R induced gating currents with similar kinetics and voltage dependence to those of m2R (Fig. 2c, d). We ensured that these were indeed gating currents as was done for m2R.

A parsimonious hypothesis would be that the receptor is modular, with separate voltage-sensing and ligand-binding domains interacting through 'coupling' domain(s) that differ in m1R and m2R. The affinity of GPCRs is universally regulated by their interaction with G proteins^{17–19}, and the voltage-dependent changes in m2R affinity are coupled to its interaction with the G protein⁶. A candidate 'coupling' domain might therefore be the third intracellular loop (L_3), which is the main G-protein-interacting part in m1R and m2R^{1,20–23}. To test this hypothesis, we employed chimaeras of m1R and m2R in which L_3 were interchanged²³, and measured ACh-induced GIRK or Cl^- currents (Supplementary Fig. 2). Indeed, L_3 of m2R conferred m2R-like

¹Department of Neurobiology, The Hebrew University, Jerusalem, 91904, Israel. ²Department of Physiology, University of Wisconsin, Madison, Wisconsin 53706, USA. ³Department of Physiology and Pharmacology, Tel Aviv University, Tel-Aviv, 69978, Israel. ⁴Institute for Molecular Pediatric Sciences, University of Chicago, Chicago, Illinois 60637, USA. [†]Present address: Neuroscience Department, The Johns Hopkins University Medical School, Baltimore, MD21205, USA.

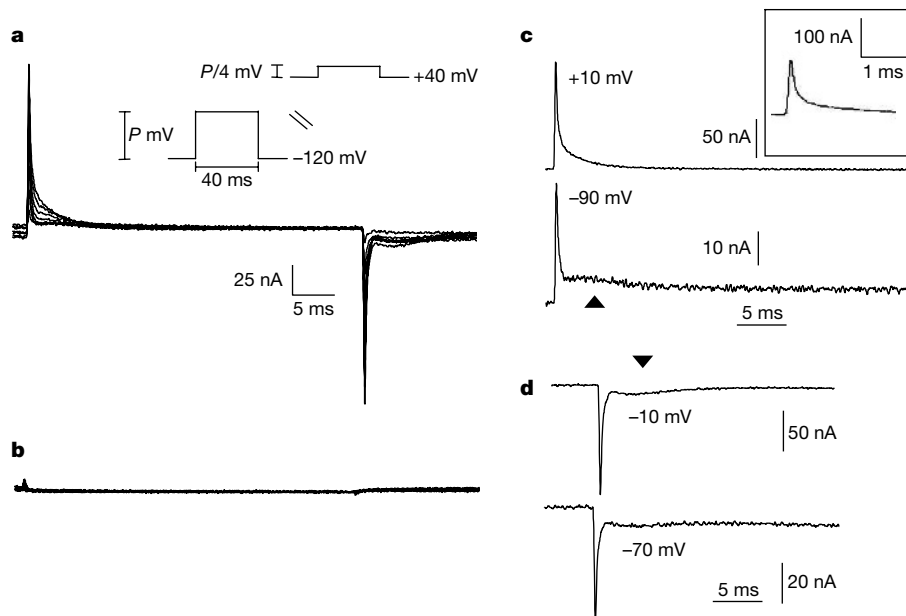


Figure 1 | Gating currents in m2R-expressing oocytes. **a**, Gating currents elicited by 40-ms depolarizing pulses (P) with 20-mV steps from a holding potential of -120 mV. Inset: the pulse protocol. Symmetric capacitive currents were subtracted by using pulses of $P/4$ mV from a holding potential of $+40$ mV. **b**, Gating currents in water-injected oocytes. **c**, Representative

ON gating currents at $+10$ or -90 mV. The arrowheads here and in **d** indicate the slow component of the current. Inset: close-up of the kinetics of the early part of the current at $+10$ mV. **d**, Representative OFF gating currents at -10 and -70 mV.

voltage dependence on m1R, and vice versa, as assayed by measuring ACh-induced GIRK or Cl^- currents (Supplementary Fig. 3) and by direct ACh binding (Supplementary Fig. 4). The G protein itself did not affect the voltage dependence, because the voltage dependence of m1R was preserved when it interacted with a carboxy-terminal mutant of $\text{G}\alpha_{i3}$ instead of G_q , activating GIRK currents (Supplementary Fig. 5).

The amino and carboxy termini of L_3 are critical for G-protein coupling^{21,22}. Noticeably, the N termini of L_3 of the $\text{G}_{i/o}$ -coupled receptors possess a string of positively charged residues that m1R lacks (Supplementary Fig. 6a), whereas the C termini show no obvi-

ous motifs of difference. We therefore replaced the N-terminal residues KKDKK (amino acids 218–222) of L_3 of m2R with the sequence ELAAL from the equivalent region of m1R. Replacement of three out of five residues did not affect the voltage dependence (Supplementary Fig. 6b). However, the replacement of the whole KKDKK motif (creating the ELAAL mutant) abolished the voltage dependence; the ELAAL mutant remained in a high-affinity state even under depolarization (Fig. 3a, and Supplementary Fig. 7, open circles) without affecting the size or shape of ACh-evoked GIRK currents (Supplementary Fig. 6c). Nevertheless, the ELAAL mutant displayed gating currents (Fig. 3b) and a $Q-V$ curve (Supplementary Fig. 7,

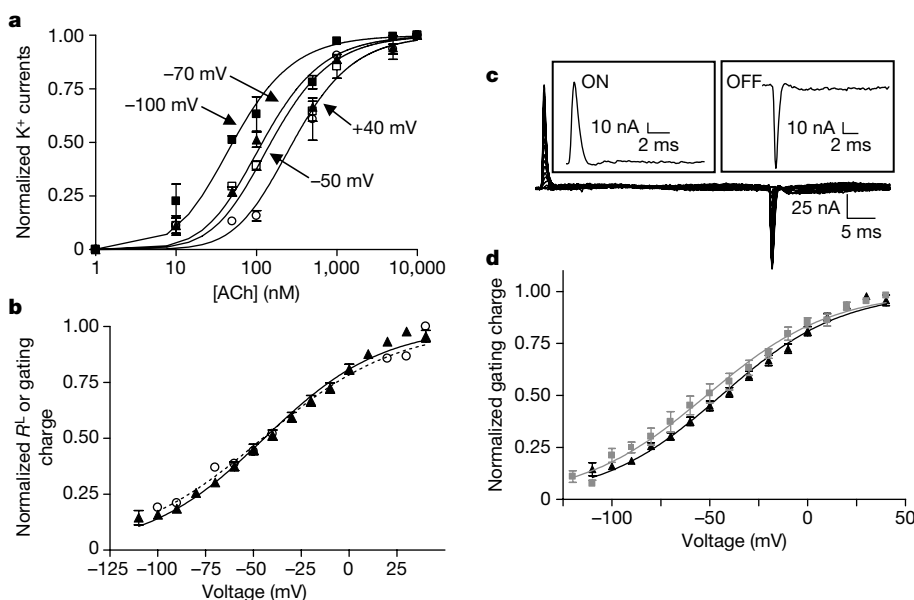


Figure 2 | Voltage dependences of gating currents of m1R and m2R, and of the agonist affinity of m2R. **a**, m2R dose-response curves at the indicated voltages. Each point shows the mean $\pm$ s.e.m. for 6–12 oocytes. Lines were fitted to equation (2) in Supplementary Methods. **b**, m2R R_1-V curve (open circles; dashed line) and $Q-V$ curve (filled triangles; solid line). Lines show

fits to the Boltzmann equation. $n = 6$ –34 for each point. **c**, Gating currents in m1R-expressing oocytes. Inset: ON gating currents at $+10$ mV and OFF gating currents at -10 mV. **d**, $Q-V$ curves (means $\pm$ s.e.m.) of m2R (black triangles; same data as in **b**) and of m1R (grey squares) ($n = 7$).

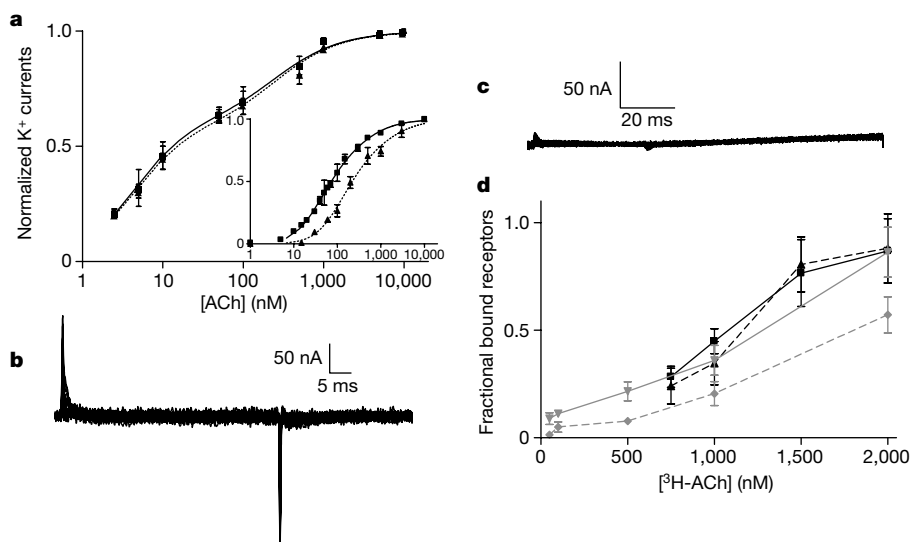


Figure 3 | Gating currents and dose-response curves of mutated m2R. **a**, Dose-response curves of m2R-ELAAL at -60 mV (squares; solid line) and $+40$ mV (triangles; dotted line). Results are given as means $\pm$ s.e.m.; $n = 8$ –21. Lines were fitted to equation (1) in Supplementary Methods. Inset: dose-response curves of m2R at the same voltages (from ref. 6, with permission). **b**, **c**, Gating currents were present in m2R-ELAAL (**b**) but not in

m2R-D120N,R121N (**c**). **d**, [^3H]ACh binding to m2R (grey; from ref. 6, with permission) and to m2R-D120N,R121N (black) at -88 mV (squares and downward triangles; solid lines) and $+5$ mV (upward triangles and diamonds; dashed lines). Lines were drawn between experimental points without fitting. Fractional specific binding was determined as described in Supplementary Methods. Results are given as means $\pm$ s.e.m.; $n = 10$ –35.

filled triangles) similar to that of native m2R. These results indicate that the KKDKK motif is not involved in voltage sensing but is essential for coupling the voltage sensor to ligand-binding affinity in m2R. The coupling of voltage sensing to binding affinity probably involves sequences besides KKDKK (or an equivalent segment in m1R), because its replacement with ELAAL did not reverse the voltage sensitivity of m2R to that of m1R.

The correlation between voltage dependences of charge movement and agonist affinity (Fig. 2b) implies, but does not prove, a causative relation between the two. A parallel perturbation of charge movement and binding affinity would provide a more direct evidence for such a relation. Unfortunately, the lack of a clear putative voltage-sensing motif (in contrast to voltage-sensitive channels) precluded a complete identification of such a sensor in this work. We attempted to hamper charge movement by focusing on the motif DRY, located in the junction between transmembrane segment 3 and intracellular loop 2, conserved in all GPCRs of family A (ref. 1), to which m1R and m2R belong. This motif includes two charged residues, Asp and Arg. Replacing its arginine residue with asparagine abolishes the m1R-induced activation of G proteins and confers a lower agonist-binding affinity^{24,25}. Because depolarization shifts m1R into a high-affinity state⁶, we conjectured that this m1R mutant is stuck in a low-affinity state because it has an impaired depolarization-induced charge movement. We therefore mutated the corresponding charged residues in m2R.

Mutation of Asp 120 to asparagine (m2R-D120N) produced a receptor with robust gating currents, normal expression in the plasma membrane, but slightly modified Q - V curve properties (Supplementary Fig. 8). Mutation of both charged residues (m2R-D120N,R121N) abolished the charge movement (Fig. 3c). The expression of this mutant, measured by quinuclidinyl benzilate (QNB) binding in intact oocytes, was about one-third of the expression of the native receptor (7 fmol per oocyte for the mutant, in contrast with 20 fmol per oocyte for native m2R). Nevertheless, it is unlikely that this low expression was the reason for the complete lack of gating currents. This is because a current that is one-third of the native m2R gating current would still be easily detectable (see Fig. 1).

In m2R, depolarization drives the receptor into a low-affinity state. The m2R-D120N,R121N mutant, which lacks gating currents, may

be expected to remain in a high-affinity state irrespective of V_m . Because this mutant failed to activate GIRK currents, we measured direct ACh binding to intact oocytes expressing the mutant (Fig. 3d). Indeed, the m2R-D120N,R121N mutant remained in a high-affinity state even under depolarization. This mutant was not deficient in its binding properties, as indicated by the maximal binding per receptor molecule, which was similar to that of the native m2R. It is therefore possible that these residues, although currently believed to be located outside the membrane electrical field, may be involved in voltage sensing.

The following model accommodates our cumulative results. GPCRs exist in two affinity states towards agonists: high affinity when coupled to the G protein, and low affinity when free^{18,19}. We suggest that depolarization-induced charge movement causes a conformational change in the loop of the GPCR that couples to the G protein. As a result, the likelihood of the receptor's coupling to the G protein is altered, and consequently the affinity of the receptor towards the agonist changes.

The finding that agonist binding of GPCRs is modulated by voltage may have a substantial physiological relevance. The regulation by voltage of this first and crucial step in GPCR activation amplifies (or attenuates) the downstream signalling. Modulation of ligand-binding affinity of GPCRs by membrane potential may be more widespread than previously suspected. Indeed, the affinities of two metabotropic glutamate receptors mGluR1 and mGluR3 are voltage sensitive²⁶. mGluR1 is involved in long-term depression in the hippocampus and in the cerebellum²⁷, and mGluR3 is involved in the modulation of glutamate release in the central nervous system²⁸. Last, one important role of the voltage-dependent modulation of affinity of GPCRs was already implied by the demonstration that a depolarization-induced shift of m2R from high to low affinity has a crucial role in determining the kinetics of ACh release from nerve terminals²⁹. The fast kinetics of the early charge movement indicate that the voltage-dependent shift in affinity may be caused by even the brief action potential, and hence might have a function in controlling transmitter release²⁹.

METHODS

Xenopus oocytes were prepared, injected with RNA, and whole-cell currents were measured with the use of a two-electrode voltage clamp as described^{6,30}. See

Supplementary Methods for details. $G\alpha_{i3}$ was coexpressed with GIRK and m2R for ion current measurements^{6,30} but not in cut-open clamp experiments. The expression of m2R alone did not increase the membrane conductance by itself and did not activate any dormant endogenous channels. The expression level of the receptor was measured with the muscarinic antagonist QNB as described⁶. **Cut-open oocyte voltage clamp.** Recordings were performed with a CA-1B amplifier (Dagan), as described¹². The external solution contained (in mM) 115 *N*-methyl-D-glucamine (NMDG)-methanesulphonate, 2 CaCl_2 , 20 HEPES, pH 7.5. The internal solution was similar but did not contain CaCl_2 and contained 2 mM EGTA. The total charge was calculated from the gating currents, assuming that with the cut-open configuration we record from one-fifth of the membrane area. Charge per receptor was obtained by dividing the total charge measured at +40 mV by the number of expressed receptors estimated from QNB binding.

Data analysis. The dose–response curves were fitted to a Michaelis–Menten type equation, assuming two affinity states and two binding sites for ACh in each, as described⁶. At each voltage, R^L was derived from the dose–response curve as explained in Supplementary Methods, and the relation between R^L and V_m was then fitted to a standard Boltzmann equation.

Statistical evaluation. Significance was checked by Student's two-tailed *t*-test.

Received 26 July; accepted 18 September 2006.

Published online 25 October 2006.

- Gether, U. Uncovering molecular mechanisms involved in activation of G protein-coupled receptors. *Endocr. Rev.* **21**, 90–113 (2000).
- Marty, A. & Tan, Y. P. The initiation of calcium release following muscarinic stimulation in rat lacrimal glands. *J. Physiol. (Lond.)* **419**, 665–687 (1989).
- Ong, B. H. *et al.* G protein modulation of voltage-sensitive muscarinic receptor signaling in mouse pancreatic acinar cells. *Pflügers Arch.* **441**, 604–610 (2001).
- Martinez-Pinna, J., Tolhurst, G., Gurung, I. S., Vandenberg, J. I. & Mahaut-Smith, M. P. Sensitivity limits for voltage control of P2Y receptor-evoked Ca^{2+} mobilization in the rat megakaryocyte. *J. Physiol. (Lond.)* **555**, 61–70 (2004).
- Martinez-Pinna, J. *et al.* Direct voltage control of signaling via P2Y1 and other G-protein-coupled receptors. *J. Biol. Chem.* **280**, 1490–1498 (2005).
- Ben-Chaim, Y., Tour, O., Dascal, N., Parnas, I. & Parnas, H. The M_2 muscarinic G-protein-coupled receptor is voltage-sensitive. *J. Biol. Chem.* **278**, 22482–22491 (2003).
- Hodgkin, A. L. & Huxley, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol. (Lond.)* **117**, 500–544 (1952).
- Armstrong, C. M. & Bezanilla, F. Currents related to movement of the gating particles of the sodium channels. *Nature* **242**, 459–461 (1973).
- Rakowski, R. F. *et al.* Charge translocation by the Na/K pump. *Ann. NY Acad. Sci.* **834**, 231–243 (1997).
- Loo, D. D., Hazama, A., Supplisson, S., Turk, E. & Wright, E. M. Relaxation kinetics of the Na^+ /glucose cotransporter. *Proc. Natl Acad. Sci. USA* **90**, 5767–5771 (1993).
- Murata, Y., Iwasaki, H., Sasaki, M., Inaba, K. & Okamura, Y. Phosphoinositide phosphatase activity coupled to an intrinsic voltage sensor. *Nature* **435**, 1239–1243 (2005).
- Stefani, E. & Bezanilla, F. Cut-open oocyte voltage-clamp technique. *Methods Enzymol.* **293**, 300–318 (1998).
- Bezanilla, F. & Armstrong, C. M. Inactivation of the sodium channel. I. Sodium current experiments. *J. Gen. Physiol.* **70**, 549–566 (1977).
- Haga, K., Haga, T. & Ichijima, A. Reconstitution of the muscarinic acetylcholine receptor. Guanine nucleotide-sensitive high affinity binding of agonists to purified muscarinic receptors reconstituted with GTP-binding proteins (G_i and G_o). *J. Biol. Chem.* **261**, 10133–10140 (1986).
- Cohen-Armon, M. & Sokolovsky, M. Evidence for involvement of the voltage-dependent Na^+ channel gating in depolarization-induced activation of G-proteins. *J. Biol. Chem.* **268**, 9824–9838 (1993).
- Ilouz, N., Branski, L., Parnis, J., Parnas, H. & Linial, M. Depolarization affects the binding properties of muscarinic acetylcholine receptors and their interaction with proteins of the exocytic apparatus. *J. Biol. Chem.* **274**, 29519–29528 (1999).
- Kenakin, T. Principles: receptor theory in pharmacology. *Trends Pharmacol. Sci.* **25**, 186–192 (2004).
- De Lean, A., Stadel, J. M. & Lefkowitz, R. J. A ternary complex model explains the agonist-specific binding properties of the adenylate cyclase-coupled β -adrenergic receptor. *J. Biol. Chem.* **255**, 7108–7117 (1980).
- Lefkowitz, R. J., Cotecchia, S., Samama, P. & Costa, T. Constitutive activity of receptors coupled to guanine nucleotide regulatory proteins. *Trends Pharmacol. Sci.* **14**, 303–307 (1993).
- Lechleiter, J. *et al.* Distinct sequence elements control the specificity of G protein activation by muscarinic acetylcholine receptor subtypes. *EMBO J.* **9**, 4381–4390 (1990).
- Wess, J. G-protein-coupled receptors: molecular mechanisms involved in receptor activation and selectivity of G-protein recognition. *FASEB J.* **11**, 346–354 (1997).
- Wess, J., Bonner, T. I., Dorje, F. & Brann, M. R. Delineation of muscarinic receptor domains conferring selectivity of coupling to guanine nucleotide-binding proteins and second messengers. *Mol. Pharmacol.* **38**, 517–523 (1990).
- Kubo, T. *et al.* Location of a region of the muscarinic acetylcholine receptor involved in selective effector coupling. *FEBS Lett.* **241**, 119–125 (1988).
- Zhu, S. Z., Wang, S. Z., Hu, J. & el-Fakahany, E. E. An arginine residue conserved in most G protein-coupled receptors is essential for the function of the m1 muscarinic receptor. *Mol. Pharmacol.* **45**, 517–523 (1994).
- Hulme, E. C., Curtis, C. A., Page, K. M. & Jones, P. G. The role of charge interactions in muscarinic agonist binding, and receptor-response coupling. *Life Sci.* **56**, 891–898 (1995).
- Ohana, L., Barchad, O., Parnas, I. & Parnas, H. The metabotropic glutamate G-protein-coupled receptors mGluR3 and mGluR1a are voltage sensitive. *J. Biol. Chem.* **281**, 24204–24215 (2006).
- Linden, D. J. Long-term synaptic depression in the mammalian brain. *Neuron* **12**, 457–472 (1994).
- Bradley, S. R. *et al.* Activation of group II metabotropic glutamate receptors inhibits synaptic excitation of the substantia nigra pars reticulata. *J. Neurosci.* **20**, 3085–3094 (2000).
- Parnas, H., Segel, L., Dudel, J. & Parnas, I. Autoreceptors, membrane potential and the regulation of transmitter release. *Trends Neurosci.* **23**, 60–68 (2000).
- Peleg, S., Varon, D., Ivanina, T., Dessauer, C. W. & Dascal, N. G. α_1 controls the gating of the G protein-activated K^+ channel, GIRK. *Neuron* **33**, 87–99 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Kubo for providing us with the clones for the m2R/L₃-m1 and m1R/L₃-m2 constructs. This work was supported by a grant to J. Dudel, I.P. and H.P. from the Deutsche Forschungsgemeinschaft, and by NIH grants to N.D., B.C. and to F.B.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.P. (hannap@huji.ac.il).

LETTERS

XIAP deficiency in humans causes an X-linked lymphoproliferative syndrome

Stéphanie Rigaud^{1*}, Marie-Claude Fondanèche^{1*}, Nathalie Lambert^{1,2}, Benoit Pasquier¹, Véronique Mateo¹, Pauline Soulas¹, Lionel Galicier³, Françoise Le Deist^{1,2}, Frédéric Rieux-Laucat¹, Patrick Revy¹, Alain Fischer^{1,4}, Geneviève de Saint Basile¹ & Sylvain Latour¹

The homeostasis of the immune response requires tight regulation of the proliferation and apoptosis of activated lymphocytes^{1,2}. In humans, defects in immune homeostasis result in lymphoproliferation disorders including autoimmunity, haemophagocytic lymphohistiocytosis and lymphomas. The X-linked lymphoproliferative syndrome (XLP) is a rare, inherited immunodeficiency that is characterized by lymphohistiocytosis, hypogammaglobulinaemia and lymphomas, and that usually develops in response to infection with Epstein–Barr virus (EBV)^{3–5}. Mutations in the signalling lymphocyte activation molecule (SLAM)-associated protein SAP, a signalling adaptor molecule, underlie 60% of cases of familial XLP^{6–8}. Here, we identify mutations in the gene that encodes the X-linked inhibitor-of-apoptosis XIAP (also termed BIRC4) in patients with XLP from three families without mutations in SAP. These mutations lead to defective expression of XIAP. We show that apoptosis of lymphocytes from XIAP-deficient patients is enhanced in response to various stimuli including the T-cell antigen receptor (TCR)–CD3 complex, the death receptor CD95 (also termed Fas or Apo-1) and the TNF-associated apoptosis-inducing ligand receptor (TRAIL-R). We also found that XIAP-deficient patients, like SAP-deficient patients, have low numbers of natural killer T-lymphocytes (NKT cells)^{9,10}, indicating that XIAP is required for the survival and/or differentiation of NKT cells. The observation that XIAP-deficiency and SAP-deficiency are both associated with a defect in NKT cells strengthens the hypothesis that NKT cells have a key role in the immune response to EBV. Furthermore, by identifying an XLP immunodeficiency that is caused by mutations in XIAP, we show that XIAP is a potent regulator of lymphocyte homeostasis *in vivo*.

We studied a cohort of 18 families with an inherited disease diagnosed as XLP. Among them, 15 were found to harbour mutations in the SAP gene (not shown). However, in affected individuals from three families (families 1, 2 and 3), we failed to detect mutations in the SAP gene (Fig. 1a). To exclude the possibility that these patients carried mutations in the regulatory regions of SAP, we analysed by western blotting the expression of SAP in the lysates from their lymphocytes (Supplementary Fig. 1). SAP was expressed normally in these individuals. The clinical phenotype of the patients without SAP mutations is similar to that seen in SAP-deficient patients: most of them (9 of 12) have developed a lymphohistiocytosis, related to EBV infection in all but two cases, and four have hypogammaglobulinaemia (Table 1). In addition, we found no gross abnormality in lymphocyte cell phenotype in these patients (Supplementary Table 1). However, these patients often had splenomegaly, which was

noticed as the first clinical manifestation of their condition, unlike SAP-deficient patients^{3,5}. These results led us to conclude that the molecular defect responsible for the XLP in these patients is not related to SAP, although the family pedigrees show that their condition is X-linked inherited (Fig. 1a).

To localize the genetic defect that underlies the condition of these patients, we performed a genotype analysis of the X chromosome in 21 individuals from families 1 and 2. A unique region of 10 cM at Xq25 that cosegregates with the condition was identified in a first screening. The mapping was then refined by genotyping additional microsatellite markers in this region, allowing us to limit the region of interest to a physical distance of 1.8 Mb (Fig. 1b). This region contains seven known genes, including SAP, and four open reading frames. By sequencing, we found that the DNA samples of the affected individuals harboured mutations in the gene encoding the X-linked inhibitor-of-apoptosis XIAP (also termed BIRC4)^{11,12} (Fig. 1c). XIAP/BIRC4 comprises six exons that encode a 56-kDa member of the family of inhibitor-of-apoptosis proteins (IAPs), which are characterized by the presence of baculovirus repeat domains (BIRs). Single nucleotide mutations consisting of a hemizygous deletion of cytidine 291 (291delC) and a hemizygous guanosine-to-thymidine transversion at position 352 (G352T) were identified in family 1 and family 3, respectively. The 291delC mutation results in a frame shift leading to a stop codon at position 387 (G99K/X129) whereas the G352T mutation causes a stop codon at position 354 (E118X). In family 2, there was a deletion of 2,606 nucleotides encompassing exon 2 of XIAP. The mothers of the patients were found to be asymptomatic heterozygous carriers and the healthy siblings did not carry mutations in XIAP. We also checked that the expression of the genes surrounding XIAP (*THO complex 2* (*THOC2*) and *Stromal antigen 2* (*STAG2*) in addition to SAP) was normal in the XIAP-mutated patients (Supplementary Fig. 2).

In normal individuals, XIAP is broadly expressed in cells from the haemopoietic system including lymphocytes, myeloid cells and natural killer (NK) cells (Supplementary Fig. 3a). In XIAP-mutated patients, the identified mutations were deleterious, as we failed to detect expression of XIAP in the lysates of lymphocytes from XIAP-mutated patients (Fig. 1d and Supplementary Fig. 3b, c). In contrast, the expression of cIAP-1 and cIAP-2, two XIAP-related proteins, was normal in XIAP-deficient patients (Fig. 1d). In cells from SAP-deficient patients, XIAP, cIAP-1 and cIAP-2 were expressed normally. Together with the normal expression of SAP in XIAP-deficient patients, these data indicate that XIAP and SAP have no influence on their reciprocal expression.

¹Inserm 768, Laboratoire du Développement Normal et Pathologique du Système Immunitaire; Univ. René Descartes, Paris, F-75015, France. ²AP-HP, Hôpital Necker Enfants-Malades, Centre d'étude des Déficiences Immunitaires, Paris, F-75015, France. ³AP-HP, Hôpital Saint-Louis, Service d'Immuno-Hématologie, Paris F-75010, France. ⁴AP-HP, Hôpital Necker Enfants-Malades, Unité d'Immunologie-Hématologie Pédiatrique, Paris F-75015 France.

*These authors contributed equally to this work.

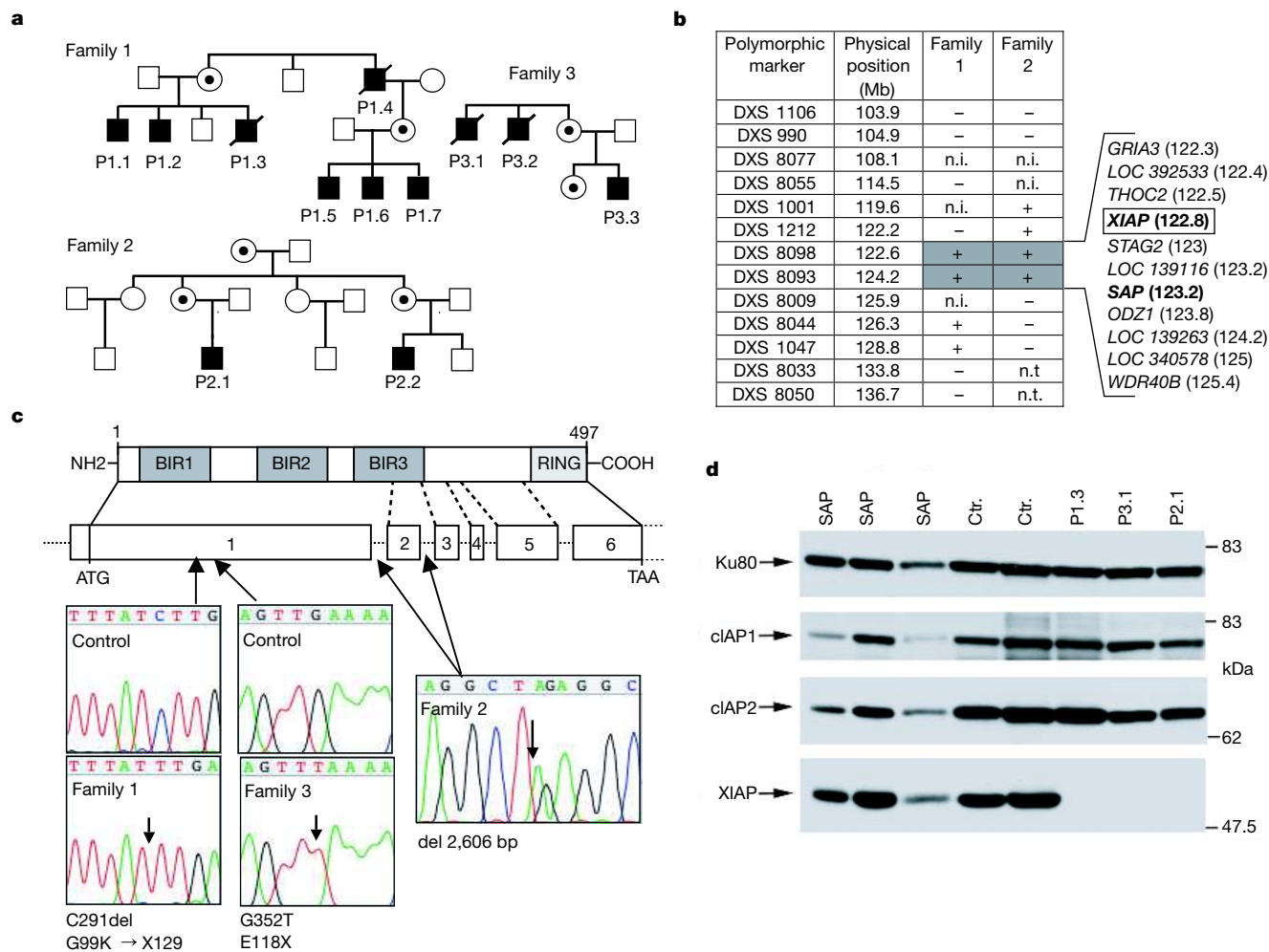


Figure 1 | Identification of mutations in XIAP in affected individuals from XLP families without SAP mutations. **a**, Pedigrees of three families presenting an XLP syndrome without detectable mutations in SAP. Black boxes represent affected individuals. Diagonal bars indicate deceased individuals. Circles with black points identify female carriers. An identification number assigned to each patient is indicated. **b**, Analysis of polymorphic markers mapped in Xq25 in families 1 and 2. The physical position of each polymorphic marker is indicated according to NCBI. (+) indicates that the polymorphic marker cosegregates with the condition in the family while (–) indicates that it does not. The different genes contained in the candidate region (grey shading) are listed and their physical position in

Mb is indicated in parentheses according to the NCBI *Homo sapiens* map. n.t., not tested; n. i., not informative. **c**, Schematic representation of the XIAP protein and the intron–exon organization of the XIAP gene. The mutations in XIAP identified in families 1, 2 and 3 are depicted by DNA electropherograms and indicated by arrows inside the electropherograms. The sites of the mutations in the XIAP gene are shown by arrows. **d**, Expression of inhibitor of apoptosis proteins (cIAP-1, cIAP-2 and XIAP) in cells from healthy control (Ctrl.), XIAP-mutated (P1.3, P2.1 and P3.1) and SAP-mutated individuals (SAP). Lysates from PHA blasts were immunoblotted with anti-XIAP, anti-cIAP-2 or anti-cIAP-1 antibodies. Immunoblotting with anti-Ku80 antibodies is shown as a loading control.

Table 1 | Clinical features of patients

Family	Origin	Patients	Age at diagnosis (years)	Age (years) and cause of death	H/S	EBV	HLH	Hypo-g	Others
1	Caucasian	P1.1	4		+	+	+	–	Colitis (4–17)
		P1.2	3		+	+	+	–	
		P1.3	1	11; HLH/BMT	+	–	+	+	
		P1.4	1.5	41; HLH	+	–	+	+	Colitis (40–41)
		P1.5	n.k		n.k	+	+	–	
		P1.6	2.5		+	+	+	–	
		P1.7	5		+	+	+	+	
2	Caucasian	P2.1	1		+	+	+	–	
		P2.2	6		+	+	+	–	
3	Caucasian	P3.1	20	20; HLH	n.k	+	+	n.k	
		P3.2	0.5	0.5; HLH	n.k	n.k	+	n.k	
		P3.3	22		+	+	–	+	

H/S, hepatosplenomegaly or splenomegaly (in brackets for all clinical features, ages at first and at latest occurrence or persistence). EBV, EBV infection (anti-EBV serology and/or EBV-positive PCR). HLH, recurrent haemophagocytic lymphohistiocytosis. HLH required treatment with corticosteroids in all patients and with cyclosporin in two patients. Hypo-g, hypogammaglobulinaemia. BMT, bone marrow transplantation. n.k., not known.

Because of the similarity of their symptoms to the clinical features of SAP deficiency, we thought that XIAP-deficient patients might share common defects. SAP deficiency results in defects in T-lymphocyte and NK cell functions that have been proposed to account for various phenotypic manifestations of XLP (ref. 4). In CD8⁺ T- and NK cells, the absence of SAP leads to a loss of function of the SLAM-related receptor 2B4, a receptor that stimulates CD8⁺- and NK- cell-mediated cytotoxicity and that has been implicated in the elimination of EBV-infected cells^{4,13}. In contrast to SAP-deficient NK cells, XIAP-deficient NK cells did not show an impairment in 2B4-mediated cytotoxicity (Supplementary Fig. 4a, b). We and others have reported that SAP is also required for the development of NKT cells^{9,10}. In SAP-deficient patients, NKT cells are absent. NKT cells represent a peculiar subpopulation of $\alpha\beta$ T cells that express an invariant TCR V α 24–V β 11, which recognizes glycosphingolipids presented by the CD1d molecule¹⁴. Examination of NKT cells by flow cytometry revealed very low numbers of NKT cells in the peripheral blood mononuclear cells (PBMCs) of XIAP-deficient patients as compared with those from age-matched healthy individuals (Fig. 2). This defect is specific to this condition because age-matched

individuals with a sporadic severe infectious mononucleosis (SIM) had normal numbers of NKT cells. The absence of XIAP selectively impaired NKT cells, because when compared to age-matched controls, the percentages of NK cells were not significantly different in XIAP-deficient patients, SAP-deficient patients and patients with SIM (Supplementary Fig. 4c). We also measured the number of NKT cells from two XIAP-mutated female carriers; these were in the lower range of controls (Supplementary Fig. 5a). To determine whether these low numbers of NKT cells could result from a loss of cells expressing the X chromosome harbouring the mutated XIAP allele, we analysed X-chromosome inactivation in one XIAP-mutated female carrier (Supplementary Fig. 5b, c). Consistent with this hypothesis, the X-chromosome inactivation was non-random within her NKT cells.

IAPs can suppress apoptosis by interacting with caspases through their BIR domains. XIAP inhibits caspases 3, 7 and 9 in response to various apoptotic stimuli including CD95 (ref. 11). Because CD95-mediated apoptosis is an essential pathway that controls lymphocyte homeostasis^{15,16}, we analysed CD95-induced apoptosis in T-cell blasts and in EBV-transformed B cells from XIAP-deficient patients. After CD95 cross-linking, apoptosis of XIAP-deficient cells was greater than apoptosis of cells from control individuals (Fig. 3a and Supplementary Figs 6a, 7a). Similarly, activation-induced cell death (AICD) following CD3 stimulation by anti-CD3 antibodies, which is known to be CD95-dependent, was markedly increased in T lymphocytes from XIAP-deficient patients¹⁶ (Fig. 3b). As expected, this enhanced cell death was inhibited by blocking anti-CD95 antibodies and anti-FasLigand antibodies (Supplementary Fig. 6b). Stimulation of the Fas-related receptor TRAIL-R also resulted in augmented apoptosis of XIAP-deficient blasts (Fig. 3c and Supplementary Fig. 6c).

These results show that lymphocytes from XIAP-deficient patients are more susceptible to apoptotic stimuli. To prove that this effect is caused by the defect in XIAP in these cells, we carried out reconstitution experiments using lentiviral infection with a plasmid containing a wild-type form of XIAP. Expression of XIAP in XIAP-deficient T-cell blasts allowed them to survive and expand selectively after AICD (Fig. 3d, e). In contrast, no such effect was seen in XIAP-deficient cells infected with an empty plasmid or in control cells infected with a XIAP-containing plasmid. Indeed, control cells express endogenous XIAP that protects them from AICD. We also analysed XIAP-deficient EBV-transformed B cells that were reconstituted with wild-type XIAP and stimulated with anti-CD95 antibodies, and the results were similar (Supplementary Fig. 7b, c). These data indicate that the increased sensitivity to cell death upon CD3 or CD95 stimulation is rescued by forced expression of XIAP in XIAP-deficient cells.

Although many studies have shown that XIAP is an inhibitor of apoptosis in cell lines, its function *in vivo* remained elusive because XIAP-deficient mice did not show abnormal apoptosis responses¹⁷. Consistent with this phenotype, we found normal numbers of NKT cells in the haemopoietic organs of these mice (unpublished data). Thus, the function of XIAP is likely to differ between mice and humans. The excess apoptosis in the absence of XIAP might account for the abnormal immune response to EBV in humans. Enhanced lymphocyte apoptosis has been shown to result in a defective immune response to acute viral infection in mice. Virus-specific CD8⁺ T cells expanded normally but disappeared owing to accelerated cell death, leading to impaired cytotoxicity and virus clearance¹⁸. Although XIAP-deficient lymphocytes are more vulnerable to cell death, XIAP-deficient patients show no T-, B- or NK-cell lymphopenia, in contrast to the almost complete absence of NKT cells. NKT cells might be particularly sensitive to apoptosis, and XIAP might be required for their survival and/or development. The fact that SAP- and XIAP-deficient patients share a defect in NKT cells emphasizes the potential role of NKT cells in the control of EBV infection. However, *in vivo* XIAP probably has a function in other leukocytes,

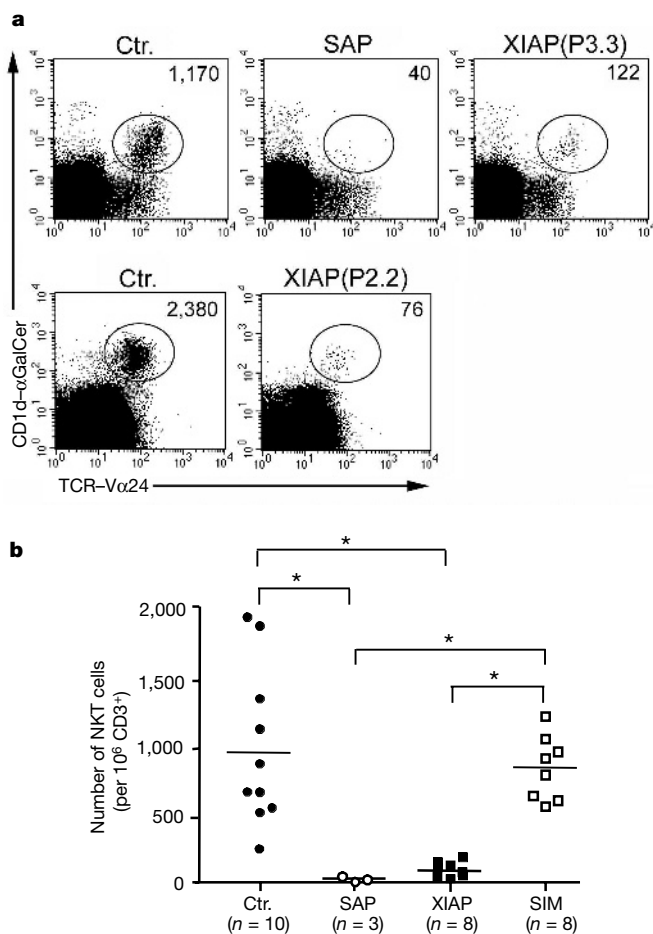


Figure 2 | Deficiency in NKT cells in XIAP-deficient patients.

a, Representative dot plots showing NKT cells in PBMCs from two healthy control (Ctr.), one SAP-deficient (SAP) and two XIAP-deficient (XIAP) individuals. Dot plots show staining with α GalCer-loaded CD1d tetramers and anti-V α 24 antibodies after gating on CD3⁺ cells. The number of NKT cells per million CD3⁺ cells in the circle gate is indicated. **b**, Comparison of the numbers of NKT cells (V α 24TCR⁺/V β 11TCR⁺) in CD3⁺ lymphocytes from control, SAP, XIAP and SIM individuals. NKT cells were analysed after gating on CD3⁺ cells by staining with anti-V α 24 and anti-V β 11 TCR antibodies. The XIAP patients tested were P1.1, P1.3, P1.7, P2.1, P2.2 and P3.3. P2.1 and P2.2 have been tested twice at an interval of six months. Bars correspond to mean values of each group of NKT cell numbers; asterisks denote $P < 0.01$.

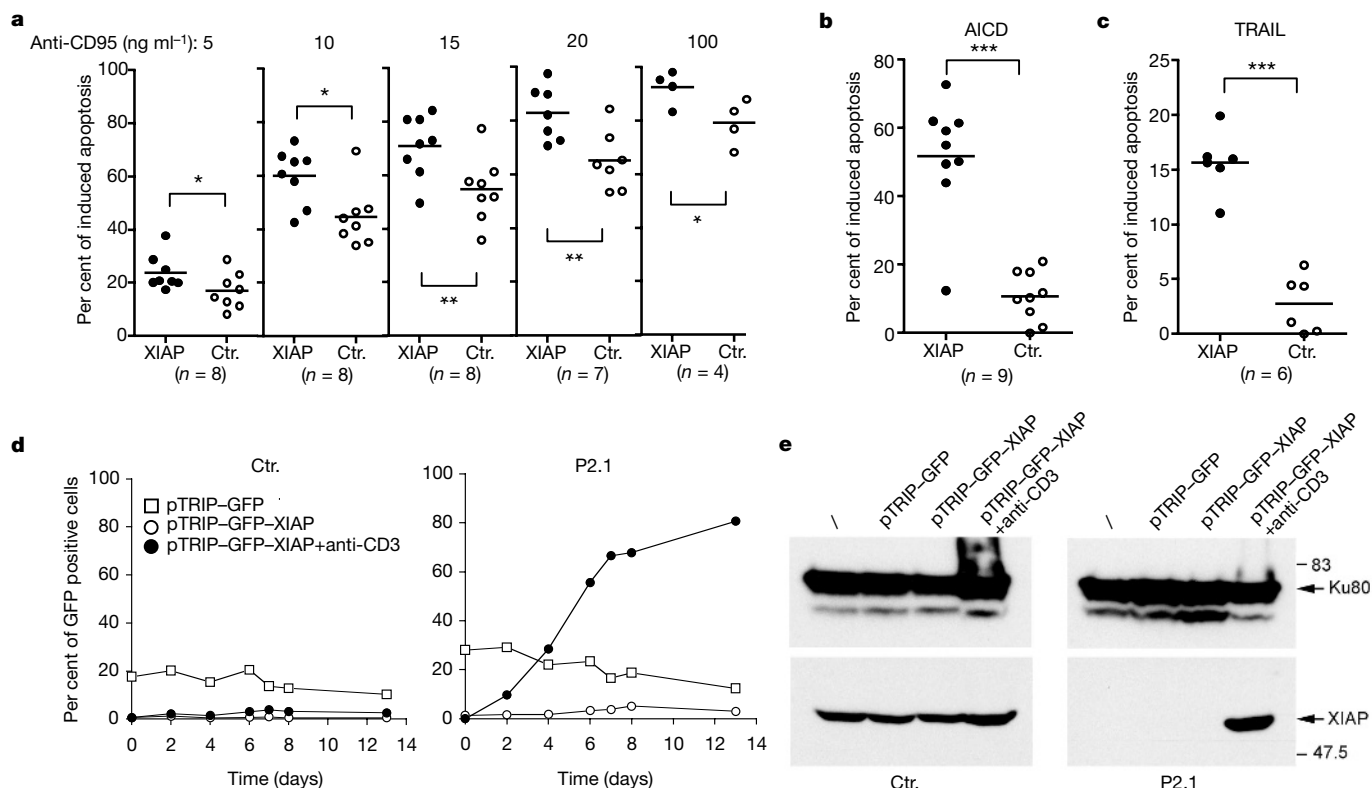


Figure 3 | Enhanced apoptosis of XIAP-deficient T lymphocytes in response to apoptotic stimuli. **a–c**, Induced apoptosis of T-cell blasts from three healthy control individuals (Ctr.) and XIAP-deficient patients (XIAP) P1.3, P2.1 and P3.1. *n* is accounted for by the fact that experiments were repeated, or not, two or three times per patient and control. Bars correspond to mean values of each group. * $P < 0.01$, ** $P < 0.001$ and *** $P < 0.0001$. **a**, Cells were stimulated with various concentrations of anti-CD95 antibodies as indicated and analysed for apoptosis. **b**, Activation-induced cell-death (AICD). Cells were stimulated with anti-CD3 antibodies and analysed for apoptosis. **c**, Induced apoptosis of cells stimulated with a trimeric TRAIL molecule. **d**, **e**, Expression of wild-type XIAP in XIAP-deficient T-cell blasts allows their selective expansion upon

activation-induced cell death. Cells from healthy control (Ctr.) and XIAP-deficient (P2.1) individuals were infected with the pTRIP-GFP-XIAP vector or the empty vector pTRIP-GFP. Cells infected with pTRIP-GFP-XIAP were then stimulated with anti-CD3 four days after infection (day 0). The percentages of GFP-positive cells were determined by flow cytometry during the culture. At day 0 before stimulation, the percentages of GFP-positive cells range from 0.76 to 1.4%. **e**, Reconstituted expression of XIAP in XIAP-deficient cells. Lysates from control (Ctr.) and XIAP-deficient (P2.1) cells at day 8 of culture, infected with pTRIP-GFP or pTRIP-GFP-XIAP and treated, or not, with anti-CD3, or non-infected (//), were immunoblotted with anti-XIAP antibodies or anti-Ku80 antibodies.

as we observed a non-random X chromosome inactivation within leukocytes of two XIAP-mutated female carriers (Supplementary Fig. 5b, c), indicating that cells expressing the wild-type XIAP allele have been selected. This observation is also supported by the survival effect of wild-type XIAP expression in XIAP-deficient T cells and EBV-transformed B cells (Fig. 3d, e and Supplementary Fig. 7). Also, defects that are not related to increased apoptosis cannot be excluded. Indeed, the lack of XIAP might also modify signalling pathways in which XIAP has been implicated, such as the transforming growth factor (TGF)- β , nuclear factor (NF) κ B, Jun N-terminal kinase (JNK), protein kinase B/Akt and copper metabolism pathways¹². Finally, our results stress the paradox that the defect in a molecule mainly characterized by its anti-apoptotic function leads to lymphoproliferative disorders. Recent studies have shown that XIAP is a potent target for the treatment of cancer, and XIAP inhibitors are being evaluated in clinical trials¹⁹. Our results indicate that inhibition of XIAP as a treatment for malignancy should be carefully considered, as it could lead to deleterious consequences.

METHODS

Genetic mapping and sequencing. Microsatellite markers were genotyped using UniSTS sequences and mapping information available from the NCBI (<http://www.ncbi.nlm.nih.gov>). The exons of XIAP and SAP were amplified from genomic DNA extracted from peripheral blood cells and sequenced using standard procedures.

Cells. Phytohaemagglutinin (PHA)-induced T-cell blasts and NK cells were obtained following standard protocols.

Immunoblots. Cell lysates and immunoblots were performed according to standard protocols. Anti-XIAP (clone 28), anti-cIAP-1 (B75-1) and anti-cIAP-2 (F30-2285) monoclonal antibodies were from BD Biosciences.

Apoptosis assay. Apoptotic cells were detected by propidium iodide staining according to standard protocols. The percentage of induced apoptosis was calculated according to the formula²⁰: $100 \times (\% \text{ of experimental apoptotic cells} - \% \text{ of spontaneously apoptotic cells}) / 100 - \% \text{ of spontaneously apoptotic cells}$.

Flow cytometry. NK and NKT cells were detected from stained PBMCs with the following monoclonal antibodies conjugated to fluorescein isothiocyanate (FITC), phycoerythrin monoclonal antibodies (PE) and allophycocyanin (APC): anti-V β 11 TCR and anti-V α 24 TCR from Beckman Coulter and anti-CD3, anti-CD56 and anti-CD16 from BD Biosciences. APC-conjugated α GalCer-loaded CD1d tetramers were kindly provided by M. Leite-de-Moraes, Hôpital Necker-Enfants Malades, Paris.

Reconstitution assay. Lentiviral infections were done with a plasmid containing the triple-stranded DNA flap of HIV-1 (pTRIP) and cDNAs for the GFP (green fluorescent protein) and the XIAP under the ubiquitin promoter (pTRIP-GFP-XIAP) as described²¹. pTRIP was a gift from P. Charneau, Institut Pasteur, Paris.

Statistical analysis. The statistical analysis of results was performed with two-tailed Student's *t*-tests.

Received 26 July; accepted 14 September 2006.

- de Saint Basile, G. & Fischer, A. Defective cytotoxic granule-mediated cell death pathway impairs T lymphocyte homeostasis. *Curr. Opin. Rheumatol.* **15**, 436–445 (2003).
- Rathmell, J. C. & Thompson, C. B. Pathways of apoptosis in lymphocyte development, homeostasis, and disease. *Cell* **109** (Suppl.), S97–107 (2002).

3. Purtilo, D. T., Cassel, C. K., Yang, J. P. & Harper, R. X-linked recessive progressive combined variable immunodeficiency (Duncan's disease). *Lancet* **1**, 935–940 (1975).
4. Nichols, K. E., Ma, C. S., Cannons, J. L., Schwartzberg, P. L. & Tangye, S. G. Molecular and cellular pathogenesis of X-linked lymphoproliferative disease. *Immunol. Rev.* **203**, 180–199 (2005).
5. Gaspar, H. B., Sharifi, R., Gilmour, K. C. & Thrasher, A. J. X-linked lymphoproliferative disease: clinical, diagnostic and molecular perspective. *Br. J. Haematol.* **119**, 585–595 (2002).
6. Nichols, K. E. *et al.* Inactivating mutations in an SH2 domain-encoding gene in X-linked lymphoproliferative syndrome. *Proc. Natl Acad. Sci. USA* **95**, 13765–13770 (1998).
7. Sayos, J. *et al.* The X-linked lymphoproliferative-disease gene product SAP regulates signals induced through the co-receptor SLAM. *Nature* **395**, 462–469 (1998).
8. Coffey, A. J. *et al.* Host response to EBV infection in X-linked lymphoproliferative disease results from mutations in an SH2-domain encoding gene. *Nature Genet.* **20**, 129–135 (1998).
9. Nichols, K. E. *et al.* Regulation of NKT cell development by SAP, the protein defective in XLP. *Nature Med.* **11**, 340–345 (2005).
10. Pasquier, B. *et al.* Defective NKT cell development in mice and humans lacking the adapter SAP, the X-linked lymphoproliferative syndrome gene product. *J. Exp. Med.* **201**, 695–701 (2005).
11. Salvesen, G. S. & Duckett, C. S. IAP proteins: blocking the road to death's door. *Nature Rev. Mol. Cell Biol.* **3**, 401–410 (2002).
12. Vaux, D. L. & Silke, J. IAPs, RINGs and ubiquitylation. *Nature Rev. Mol. Cell Biol.* **6**, 287–297 (2005).
13. Parolini, S. *et al.* X-linked lymphoproliferative disease. 2B4 molecules displaying inhibitory rather than activating function are responsible for the inability of natural killer cells to kill Epstein-Barr virus-infected cells. *J. Exp. Med.* **192**, 337–346 (2000).
14. Taniguchi, M., Harada, M., Kojo, S., Nakayama, T. & Wakao, H. The regulatory role of V α 14 NKT cells in innate and acquired immune response. *Annu. Rev. Immunol.* **21**, 483–513 (2003).
15. Rieux-Laucat, F., Fischer, A. & Deist, F. L. Cell-death signaling and human disease. *Curr. Opin. Immunol.* **15**, 325–331 (2003).
16. Lenardo, M. *et al.* Mature T lymphocyte apoptosis—immune regulation in a dynamic and unpredictable antigenic environment. *Annu. Rev. Immunol.* **17**, 221–253 (1999).
17. Harlin, H., Reffey, S. B., Duckett, C. S., Lindsten, T. & Thompson, C. B. Characterization of XIAP-deficient mice. *Mol. Cell. Biol.* **21**, 3604–3608 (2001).
18. Wasem, C. *et al.* Sensitizing antigen-specific CD8⁺ T cells for accelerated suicide causes immune incompetence. *J. Clin. Invest.* **111**, 1191–1199 (2003).
19. Schimmer, A. D., Dalili, S., Batey, R. A. & Riedl, S. J. Targeting XIAP for the treatment of malignancy. *Cell Death Differ.* **13**, 179–188 (2006).
20. Strauss, G., Knape, I., Melzner, I. & Debatin, K. M. Constitutive caspase activation and impaired death-inducing signaling complex formation in CD95-resistant, long-term activated, antigen-specific T cells. *J. Immunol.* **171**, 1172–1182 (2003).
21. Zennou, V. *et al.* HIV-1 genome nuclear import is mediated by a central DNA flap. *Cell* **101**, 173–185 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the patients and their families that participated in this study, E. Le Gall (Centre Hospitalier Universitaire de Rennes) and M. Hamidou (Centre Hospitalier Universitaire de Nantes) for the blood samples and their contribution to the follow up of patients from family 1 and 3; A. Veillette and C. Picard for discussions and critical reading of the manuscript; and P. Charneau, M. Bonneville and M. Leite-de-Moraes for their gift of reagents. We also acknowledge the contribution of C. Charon from the Centre National de Génotypage (CNG, Evry, France). This work was supported by grants from the Institut National de la Santé et de la Recherche Médicale (France), the Agence National de la Recherche (France), the Association pour la Recherche contre le Cancer (France), GIS-Institut des Maladies Rares (France) and La Ligue contre le Cancer (France). B.P. was supported by a fellowship from La Ligue contre le Cancer and V.M. by a European Molecular Biology Organization fellowship. S.L. and P.R. are scientists from the Centre National pour la Recherche Scientifique (France).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.L. (latour@necker.fr).

Surface expression of MHC class II in dendritic cells is controlled by regulated ubiquitination

Jeoung-Sook Shin¹, Melanie Ebersold¹, Marc Pypaert¹, Lelia Delamarre¹, Adam Hartley¹ & Ira Mellman¹

Dendritic cells have a unique function in the immune response owing to their ability to stimulate immunologically naive T lymphocytes¹. In response to microbial and inflammatory stimuli, dendritic cells enhance their capacity for antigen presentation by a process of terminal differentiation, termed maturation^{2,3}. The conversion of immature to mature dendritic cells is accompanied by a marked cellular reorganization, including the redistribution of major histocompatibility complex class II molecules (MHC II) from late endosomal and lysosomal compartments to the plasma membrane^{4–7} and the downregulation of some forms of endocytosis, which has been thought to slow the clearance of MHC II from the surface^{8–11}. The relative extent to which these or other mechanisms contribute to the regulation of surface MHC II remains unclear, however. Here we find that the MHC II β -chain cytoplasmic tail is ubiquitinated in mouse immature dendritic cells. Although only partly required for the sequestration of MHC II in multivesicular bodies, this modification is essential for endocytosis. Notably, ubiquitination of MHC II ceased upon maturation, resulting in the accumulation of MHC II at the cell surface. Dendritic cells thus exhibit a unique ability to regulate MHC II surface expression by selectively controlling MHC II ubiquitination.

Covalent attachment of ubiquitin polymers to lysine residues of a target protein has long been known to trigger proteasome-dependent degradation. However, ubiquitination also has a function in controlling membrane transport by facilitating endocytosis and lysosomal degradation^{12,13}. Many receptor protein tyrosine kinases are transiently mono- or oligo-ubiquitinated after ligand binding, which leads to internalization and inclusion in the internal vesicles of multivesicular bodies (MVBs) concomitant with receptor de-ubiquitination^{14,15}. MHC II molecules are localized in MVBs in immature dendritic cells, although neither their mechanism of entry nor their fate is known.

Recently, a conserved lysine on the MHC II β -chain (Fig. 1a) was shown to be ubiquitinated when MHC II was transfected into 293 cells overexpressing the E3 ligase c-MIR, raising the possibility that ubiquitination might be involved in regulating MHC II function in dendritic cells¹⁶. To determine whether endogenous β -chain was ubiquitinated in dendritic cells under physiological conditions, we immunoprecipitated MHC II molecules from bone-marrow-derived and splenic dendritic cells, blotting the precipitates using an anti-ubiquitin antibody. A ladder of proteins was detected, starting at ~40 kDa and increasing in molecular weight by ~10 kDa each (Fig. 1b, c), consistent with oligo-ubiquitination of the 31-kDa MHC II β -chain by 1–5 ubiquitin monomers. To confirm their identity, we used dendritic cells from MHC II β -EGFP knockin mice. The ubiquitinated species increased by ~26 kDa, the molecular weight of green fluorescent protein (GFP) (Fig. 1d). Retroviral transduction

was next used to express, in dendritic cells cultured from MHC II β -deficient mice, wild-type MHC II β or a MHC II β (K>R) mutant, where the single lysine (K) of the β -chain cytoplasmic domain is replaced by arginine (R). Although the expression levels were lower, the transduced wild-type β -chain revealed the same oligo-ubiquitin ladder, whereas the MHC II β (K>R) mutant did not (Fig. 1e).

We next used retroviral expression to investigate whether ubiquitination affected the distribution of MHC II. To identify infected cells, we inserted an IRES (internal ribosome entry site)-EGFP complementary DNA downstream of the MHC II β sequence. Wild-type MHC II was, as expected, localized predominantly to H2-M⁺ late endosomes and lysosomes in immature dendritic cells (Fig. 2a, upper panels). In contrast, MHC II β (K>R) was found primarily at the plasma membrane (Fig. 2a, lower panels). These results were quantified by fluorescence-activated cell sorting (FACS) analysis (Fig. 2b, upper panels). We selectively analysed dendritic cell populations by gating on cells positive for both CD11c and GFP. MHC II β (K>R)-expressing dendritic cells (blue) exhibited a >20-fold increase in surface levels of MHC II compared with wild type (red) (Fig. 2b, lower-left panel). Thus, an inability to ubiquitinate MHC II correlates with a high level of surface expression in immature dendritic cells.

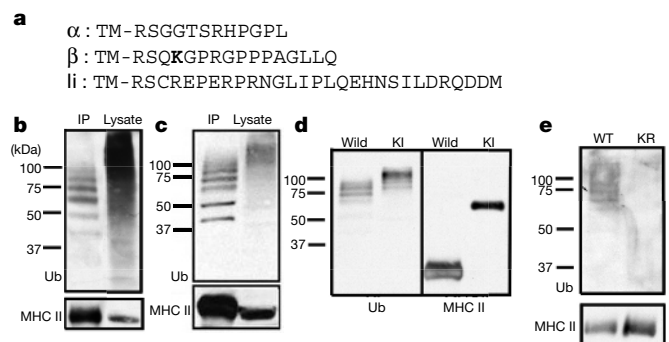


Figure 1 | MHC class II molecules are ubiquitinated in dendritic cells. **a**, The amino acid sequence of the cytosolic tail of MHC II (IA^b) α -chain, MHC II (IA^b) β -chain and invariant chain (Ii). Note a single lysine (bold K) in the β -chain. TM, transmembrane domain. **b–e**, Western blots of dendritic cell MHC II immunoprecipitates (IP) using anti-ubiquitin (Ub) antibody and anti-MHC II antibodies. Shown are: **b**, BMDCs cultured from C57/BL6 mice; **c**, splenic dendritic cells isolated from C57/BL6 mice; **d**, BMDCs cultured from C57/BL6 mice (wild) versus MHC II β -EGFP knockin mice (KI); **e**, BMDCs expressing MHC II β wild-type (WT) versus MHC II β (K>R) mutant (KR). These experiments were performed at least three times independently, and representative data are shown.

¹Department of Cell Biology and Section of Immunobiology, Ludwig Institute for Cancer Research, Yale University School of Medicine, 333 Cedar Street, PO Box 208002, New Haven, Connecticut 06520-8002, USA.

Immunoelectron microscopy revealed that MVBs were well labelled by MHC II-specific gold particles in cells expressing wild-type MHC II (Fig. 2c, left). MVBs in MHC II β (K>R)-expressing dendritic cells, on the other hand, were labelled by many fewer gold particles; instead, large numbers of gold particles decorated the plasma membrane (Fig. 2c, right). The reduced amount of MHC II β (K>R) in MVBs might reflect a decrease in biosynthetic or endocytic delivery to late endosomal compartments, or decreased sequestration in the MVBs' internal vesicles. To explore the latter possibility, we counted the number of gold particles associated with the limiting membranes versus the internal vesicles, defined objectively as particles >20 nm from the limiting membrane. The quantitative analysis (Fig. 2c) showed that a significantly lower fraction of gold particles (31%) was found in the interior of MVBs when mutant MHC II was expressed compared with wild type (51%). Thus, although ubiquitination facilitates MHC II sorting into the internal vesicles of MVBs, it may not be an absolute requirement. In any event, the most notable feature of MHC II β (K>R)-expressing dendritic cells was the high level of surface MHC II in immature cells. Conceivably, defective endocytosis of mutant MHC II might produce this phenotype if ubiquitination is required for internalization. Although it has long been known that MHC II $\alpha\beta$ dimers are internalized^{17–21}, the nature of the endocytosis signal has remained unclear. Dendritic cells expressing wild-type or MHC II β (K>R) were incubated with anti-MHC II monoclonal antibody and imaged by confocal microscopy. In wild-type MHC II-expressing dendritic cells, MHC II-monoclonal antibody complexes were found in intracellular compartments, some of which co-localized with H2-M (Fig. 2d). In contrast, little intracellular monoclonal antibody was detected in cells

expressing MHC II β (K>R); rather it appeared to be restricted to the plasma membrane (Fig. 2d). Quantification of 20 confocal images (see Supplementary Methods) confirmed a significant and consistent difference: MHC II β (K>R)-expressing dendritic cells had ~2.5-fold less fluorescence in the interior (Fig. 2e). Similar results were obtained using IgG or monovalent Fab fragments of the anti-MHC II monoclonal antibody. These data strongly suggest that ubiquitination has an important role in the endocytosis of MHC II.

We next tested whether ubiquitination is involved in MHC II degradation. Western blot analysis revealed that the total amount of MHC II in the CD11c⁺ population (dendritic cells) was about fourfold greater in cells expressing MHC II β (K>R) as opposed to wild-type MHC II (Fig. 3a). This did not reflect a difference in retroviral transduction efficiency because both cell populations contained comparable amounts of GFP. Moreover, wild-type and mutant MHC II were at similar levels in the CD11c[−] population (non-dendritic cells) (Fig. 3a).

Similar results were obtained by immunoprecipitation (Fig. 3b). Notably, the increased amount of MHC II β (K>R) was largely in the SDS stable conformation, suggesting that it was peptide-bound (Fig. 3b, arrow). Similarly, at least a small fraction of both the mutant and wild-type MHC II could be loaded with peptide after exposure of immature dendritic cells to exogenous recombinant I-E α chain (Supplementary Fig. 1). Both cell types also contained similar amounts of the MHC II-invariant (Ii) chain complex (Fig. 3b). Thus, the increased amount of MHC II β (K>R) on the cell surface was present as free or peptide-loaded $\alpha\beta$ dimers, with the Ii chain having been removed on the way to the surface. These results also emphasize that immature dendritic cells are capable of some level of

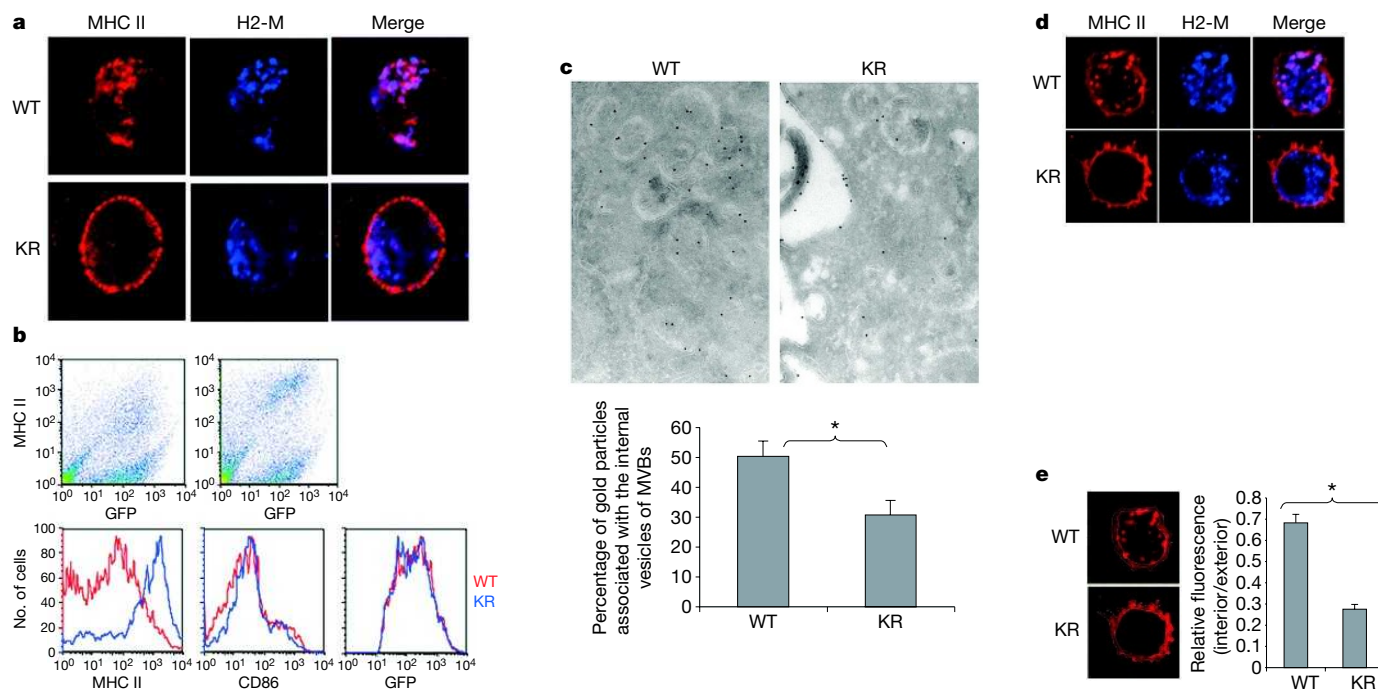


Figure 2 | Differential localization and endocytosis of wild-type MHC II versus the ubiquitin-deficient MHC II β (K>R) mutant. **a**, Immunofluorescence microscopy of BMDCs expressing wild-type MHC II (WT) and the MHC II β (K>R) mutant (KR). **b**, FACS analysis of BMDCs expressing wild-type MHC II or MHC II β (K>R). Upper panels show the surface level of MHC II in the total cell population; lower panels show the expression of designated molecules after selective gating for CD11c⁺GFP⁺ cells.

c, Quantitative immunoelectron microscopy of BMDCs expressing wild-type MHC II or MHC II β (K>R). MHC II was labelled with gold particles of 10 nm diameter. A representative micrograph for each is shown. The bar graph shows the mean and s.e.m. ($n = 46$ for wild type, $n = 70$ for MHC

II β (K>R); asterisk, $P < 0.01$, Student's t -test). **d**, Endocytosis of MHC II in BMDCs expressing wild-type MHC II or MHC II β (K>R). Cells were incubated with the anti-MHC II monoclonal antibody for 1 h at 37 °C, then stained for the MHC II antibody (red) and H2-M (blue). **e**, Quantification of MHC II antibody internalization. The interior and exterior of the z -axis sections at 2.4 μ m from the coverslip were manually marked as shown, and the mean MHC II antibody fluorescence intensity was quantified in each area; the relative intensity is shown in the bar graph where data are expressed as the mean and s.e.m. ($n = 20$ for both wild type and MHC II β (K>R); asterisk, $P < 0.001$, Student's t -test).

peptide–MHC II production, at least for select antigens²², although both the production and accumulation of peptide-loaded complexes are greatly enhanced by maturation^{3,4}.

Degradation of wild-type MHC II appeared to reflect lysosomal proteolysis. Treatment of dendritic cells with the proteasome inhibitor epoxomicin barely changed MHC II levels, although there was a dose-dependent increase of high-molecular-weight poly-ubiquitinated proteins (Fig. 3c). In contrast, the lysosomotropic agent chloroquine caused a significant increase in the amount of both MHC II and Ii chain, where the Ii chain is known to be degraded by lysosomal proteases (Fig. 3c). Because the initial rate of synthesis of both wild-type and mutant MHC II was comparable (pulse labelling with ³⁵S-methionine for 15–60 min; not shown), we conclude that ubiquitination enhances the rate of MHC II degradation in lysosomes, thus decreasing its accumulation at equilibrium.

It was intriguing that the predominant localization, and decreased endocytosis, of the MHC IIβ(K>R) mutant at the plasma membranes of immature dendritic cells resembled the distribution of wild-type MHC II in mature dendritic cells. We therefore wondered whether MHC II ubiquitination might cease upon dendritic cell maturation. Notably, MHC II immunoprecipitated from mature bone-marrow-derived dendritic cells (BMDCs) contained no detectable ubiquitin (Fig. 4a). The amount of total cell (poly-)ubiquitinated proteins did not decrease, but actually increased slightly upon maturation. We confirmed these observations for dendritic cells *in vivo*, by analysing splenic dendritic cells matured by injecting

mice with lipopolysaccharide (LPS) 16 h before isolation (Fig. 4b). Therefore, dendritic cell maturation negatively and selectively regulates the ubiquitination of MHC II.

We next performed a gain-of-function experiment to determine whether the behaviour of MHC II in mature dendritic cells could be reverted to that observed in immature dendritic cells by synthetically ubiquitinating MHC II molecules in mature dendritic cells. A cDNA encoding a single ubiquitin was fused to the cytoplasmic tail of MHC IIβ(K>R) to yield a constitutively ubiquitinated MHC II molecule (MHC IIβ(K>R)Ub). Unlike the non-ubiquitinated MHC IIβ(K>R) mutant, but similarly to wild-type MHC II, the MHC IIβ(K>R)Ub mutant was primarily localized in the lysosomes of immature dendritic cells (Fig. 4c, upper panels). Notably, MHC IIβ(K>R)Ub retained its lysosomal localization pattern in mature dendritic cells after LPS treatment (Fig. 4c, lower panels). Indeed, surface expression of the MHC IIβ(K>R)Ub mutant barely increased after LPS-induced maturation (Fig. 4d).

Finally, we assayed the endocytosis of the MHC IIβ(K>R)Ub mutant in immature and mature dendritic cells. In immature

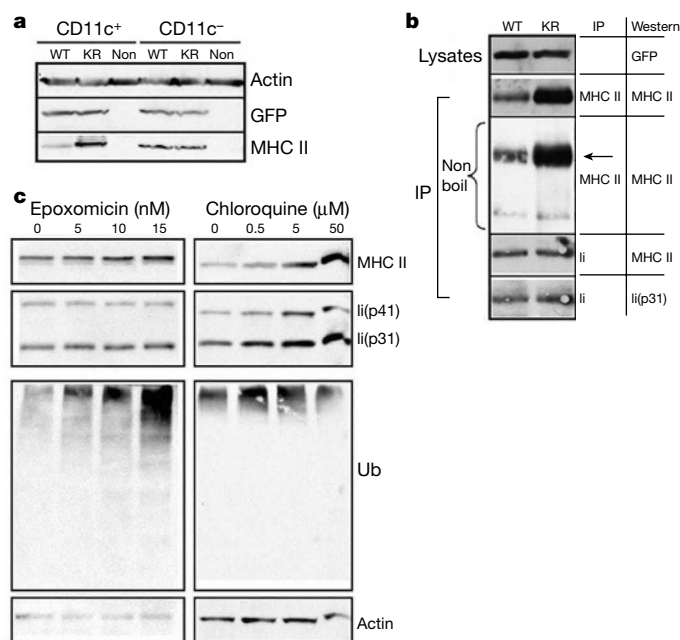


Figure 3 | Lysosomal degradation and Ii chain association of ubiquitinated and non-ubiquitinated MHC II in BMDCs. **a**, Western blot of total lysates of CD11c⁺ and CD11c⁻ populations of BMDC cultures retrovirally transduced to express wild-type MHC II (WT) or MHC IIβ(K>R) mutant. Non-transduced cells (Non) were also loaded as a control. These experiments were performed twice independently and representative data are shown. **b**, Increased levels of MHC II in dendritic cells expressing MHC IIβ(K>R) reflect the accumulation of MHC II αβ dimers and not αβ–Ii chain complexes. The whole-cell lysates and immunoprecipitates from wild-type or MHC IIβ(K>R)-expressing dendritic cells were analysed by western blotting. The antibodies used for western blotting or immunoprecipitation are listed on the right. The arrow indicates SDS-stable MHC II αβ dimers. These experiments were performed at least twice independently and representative data are shown. **c**, Lysosomal but not proteasomal inhibitors increase the overall levels of endogenous MHC II and Ii chain in wild-type BMDCs. Western blots were performed using whole lysates prepared from BMDCs treated with each drug. These experiments were performed three times independently and representative data are shown.

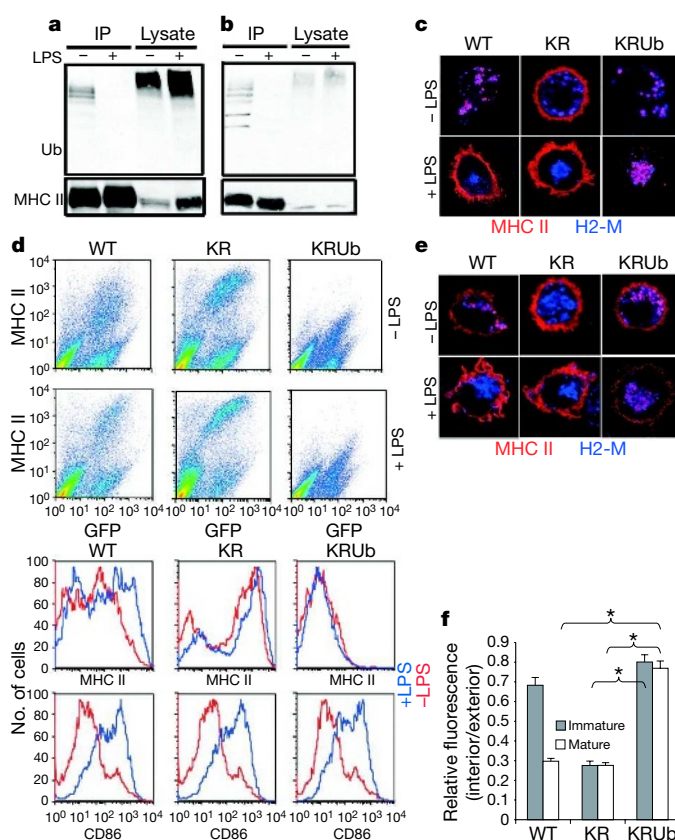


Figure 4 | Loss of MHC II ubiquitination in mature dendritic cells and its role in regulating endocytosis. **a**, **b**, Immunoprecipitation of MHC II from BMDCs (**a**) and splenic dendritic cells (**b**) that are immature (–LPS) or mature (+LPS). LPS-induced maturation of BMDCs and splenic dendritic cells was confirmed by FACS analysis (Supplementary Fig. 3). These experiments were performed at least three times independently and representative data are shown. **c**, Immunofluorescence microscopy of immature (–LPS) or mature (+LPS) BMDCs expressing wild-type MHC II (WT), MHC IIβ(K>R) mutant (KR), or MHC IIβ(K>R)Ub mutant (KRUb). **d**, FACS profiles of immature (–LPS) and mature (+LPS) BMDCs expressing wild-type, MHC IIβ(K>R), or MHC IIβ(K>R)Ub. The upper two rows show the surface level of MHC II in the total cell population; the lower two rows show the level of designated molecules after selective gating for CD11c⁺GFP⁺ cells. **e**, **f**, Endocytosis of MHC II in immature (–LPS) or mature (+LPS) BMDCs expressing wild-type, MHC IIβ(K>R), or MHC IIβ(K>R)Ub. The same assay and method of quantification as described for Fig. 2d, e were used. Data are expressed as mean and s.e.m. ($n = 20$ for each; asterisk, $P < 0.001$, Student's t -test).

dendritic cells, the MHC II β (K>R)Ub mutant internalized anti-MHC II monoclonal antibody (IgG or Fab) comparably to wild-type MHC II, both of which were clearly more efficient than the non-ubiquitinated MHC II β (K>R) mutant (Fig. 4e, upper panels). In mature dendritic cells, however, only the MHC II β (K>R)Ub mutant mediated antibody internalization and delivery to late endocytic compartments (Fig. 4e, lower panels). The continued efficient internalization by the MHC II β (K>R)Ub mutant was confirmed by quantification (Fig. 4f). This result is particularly notable given that cells expressing the MHC II β (K>R)Ub mutant expressed relatively little surface MHC II and thus bound relatively little antibody as compared with the wild type or the MHC II β (K>R) mutant. Ubiquitination is therefore required for efficient MHC II endocytosis in both immature and mature dendritic cells, indicating that it is the selective downregulation of MHC II ubiquitination, and not the cessation of macropinocytosis, that normally enables the retention of MHC II at the cell surface.

Unlike many surface receptors for which ubiquitination and endocytosis is regulated by ligand binding, MHC II ubiquitination is regulated by dendritic cell maturation. This newly discovered mode of regulation clarifies how MHC II–peptide complexes remain at the plasma membrane after their translocation from the lysosomal compartment in mature dendritic cells, despite the continued formation of clathrin-coated vesicles⁸. It also illustrates that the control of endocytosis has an important role in regulating the distribution of MHC II in dendritic cells. The fraction of newly synthesized MHC II molecules that reaches the plasma membrane on the way to lysosomes may vary in different dendritic cell types^{2,4}. However, even if this fraction is small it might still lead to large amounts of surface MHC II in immature dendritic cells at equilibrium, if MHC II was not subject to efficient internalization. As is the case for receptor tyrosine kinases^{23–25}, only a small fraction of MHC II molecules is ubiquitinated in immature dendritic cells at any one time, suggesting that ubiquitin addition to MHC II is transient, with ubiquitin moieties removed after internalization or MVB delivery. Indeed, ubiquitinated MHC II cannot mediate uptake of non-ubiquitinated $\alpha\beta$ dimers (Supplementary Fig. 2), suggesting that the entire cohort must receive the modification at some point.

In principle, ubiquitination could regulate MHC II transport in dendritic cells at two other sites: exit from the Golgi complex and sorting into MVBs. Although a role for ubiquitination in the Golgi complex has been observed in other systems^{26,27}, it seems unlikely to have a role here. Dendritic cells expressing the MHC II β (K>R) mutant did not exhibit an increase of MHC II–Ii chain complexes, suggesting that biosynthetic transport of MHC II–Ii from the Golgi to endosomes, the site of Ii chain removal and possibly peptide loading^{28,29}, occurs in the absence of ubiquitin. Sorting into MVBs, however, is partly but incompletely reduced in the MHC II β (K>R) mutant. As such, MHC II may be one of the first examples of a ubiquitin-independent cargo during MVB biogenesis. It will be important to examine this issue in detail, and also to establish where ubiquitination of MHC II occurs and how it is regulated.

METHODS

Details of cells, antibodies, plasmids, retroviral transduction, immunoprecipitation, MHC class II endocytosis and quantification, electron microscopy and quantification, and I-E α peptide/MHC class II complex expression are given in Supplementary Information.

Received 16 June; accepted 18 September 2006.

Published online 8 October 2006.

1. Banchereau, J. & Steinman, R. M. Dendritic cells and the control of immunity. *Nature* **392**, 245–252 (1998).
2. Mellman, I. & Steinman, R. M. Dendritic cells: specialized and regulated antigen processing machines. *Cell* **106**, 255–258 (2001).
3. Reis e Sousa, C. Dendritic cells in a mature age. *Nature Rev. Immunol.* **6**, 476–483 (2006).
4. Trombetta, E. S. & Mellman, I. Cell biology of antigen processing *in vitro* and *in vivo*. *Annu. Rev. Immunol.* **23**, 975–1028 (2005).

5. Chow, A., Toomre, D., Garrett, W. & Mellman, I. Dendritic cell maturation triggers retrograde MHC class II transport from lysosomes to the plasma membrane. *Nature* **418**, 988–994 (2002).
6. Turley, S. J. *et al.* Transport of peptide-MHC class II complexes in developing dendritic cells. *Science* **288**, 522–527 (2000).
7. Boes, M. *et al.* T-cell engagement of dendritic cells rapidly rearranges MHC class II transport. *Nature* **418**, 983–988 (2002).
8. Garrett, W. S. *et al.* Developmental control of endocytosis in dendritic cells by Cdc42. *Cell* **102**, 325–334 (2000).
9. Wilson, N. S., El-Sukkari, D. & Villadangos, J. A. Dendritic cells constitutively present self antigens in their immature state *in vivo* and regulate antigen presentation by controlling the rates of MHC class II synthesis and endocytosis. *Blood* **103**, 2187–2195 (2004).
10. West, M. A., Prescott, A. R., Eskelinen, E. L., Ridley, A. J. & Watts, C. Rac is required for constitutive macropinocytosis by dendritic cells but does not control its downregulation. *Curr. Biol.* **10**, 839–848 (2000).
11. Sallusto, F., Cella, M., Danieli, C. & Lanzavecchia, A. Dendritic cells use macropinocytosis and the mannose receptor to concentrate macromolecules in the major histocompatibility complex class II compartment: downregulation by cytokines and bacterial products. *J. Exp. Med.* **182**, 389–400 (1995).
12. Hicke, L. & Dunn, R. Regulation of membrane protein transport by ubiquitin and ubiquitin-binding proteins. *Annu. Rev. Cell Dev. Biol.* **19**, 141–172 (2003).
13. Raiborg, C., Rusten, T. E. & Stenmark, H. Protein sorting into multivesicular endosomes. *Curr. Opin. Cell Biol.* **15**, 446–455 (2003).
14. Marmor, M. D. & Yarden, Y. Role of protein ubiquitylation in regulating endocytosis of receptor tyrosine kinases. *Oncogene* **23**, 2057–2070 (2004).
15. Katzmann, D. J., Odorizzi, G. & Emr, S. D. Receptor downregulation and multivesicular-body sorting. *Nature Rev. Mol. Cell Biol.* **3**, 893–905 (2002).
16. Ohmura-Hoshino, M. *et al.* Inhibition of MHC class II expression and immune responses by c-MIR. *J. Immunol.* **177**, 341–354 (2006).
17. Pinet, V., Vergelli, M., Martin, R., Bakke, O. & Long, E. O. Antigen presentation mediated by recycling of surface HLA-DR molecules. *Nature* **375**, 603–606 (1995).
18. Cella, M., Engering, A., Pinet, V., Pieters, J. & Lanzavecchia, A. Inflammatory stimuli induce accumulation of MHC class II complexes on dendritic cells. *Nature* **388**, 782–787 (1997).
19. Zhong, G., Romagnoli, P. & Germain, R. N. Related leucine-based cytoplasmic targeting signals in invariant chain and major histocompatibility complex class II molecules control endocytic presentation of distinct determinants in a single protein. *J. Exp. Med.* **185**, 429–438 (1997).
20. Pathak, S. S., Lich, J. D. & Blum, J. S. Cutting edge: editing of recycling class II:peptide complexes by HLA-DM. *J. Immunol.* **167**, 632–635 (2001).
21. Salamero, J., Humbert, M., Cosson, P. & Davoust, J. Mouse B lymphocyte specific endocytosis and recycling of MHC class II molecules. *EMBO J.* **9**, 3489–3496 (1990).
22. Colledge, L., Bennett, C. L., Reay, P. A. & Blackburn, C. C. Rapid constitutive generation of a specific peptide-MHC class II complex from intact exogenous protein in immature murine dendritic cells. *Eur. J. Immunol.* **32**, 3246–3255 (2002).
23. Mosesson, Y. *et al.* Endocytosis of receptor tyrosine kinases is driven by monoubiquitylation, not polyubiquitylation. *J. Biol. Chem.* **278**, 21323–21326 (2003).
24. Miyake, S., Lupher, M. L. Jr, Druker, B. & Band, H. The tyrosine kinase regulator Cbl enhances the ubiquitination and degradation of the platelet-derived growth factor receptor α . *Proc. Natl Acad. Sci. USA* **95**, 7927–7932 (1998).
25. Haglund, K. *et al.* Multiple monoubiquitination of RTKs is sufficient for their endocytosis and degradation. *Nature Cell Biol.* **5**, 461–466 (2003).
26. Beck, T., Schmidt, A. & Hall, M. N. Starvation induces vacuolar targeting and degradation of the tryptophan permease in yeast. *J. Cell Biol.* **146**, 1227–1238 (1999).
27. Helliwell, S. B., Losko, S. & Kaiser, C. A. Components of a ubiquitin ligase complex specify polyubiquitination and intracellular trafficking of the general amino acid permease. *J. Cell Biol.* **153**, 649–662 (2001).
28. Villadangos, J. A., Riese, R. J., Peters, C., Chapman, H. A. & Ploegh, H. L. Degradation of mouse invariant chain: roles of cathepsins S and D and the influence of major histocompatibility complex polymorphism. *J. Exp. Med.* **186**, 549–560 (1997).
29. Matza, D., Kerem, A. & Shachar, I. Invariant chain, a chain of command. *Trends Immunol.* **24**, 264–268 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Mellman laboratory group for advice and support. J.-S.S. is a Fellow of the Jane Coffin Childs Foundation. I.M. is an Affiliate Member of the Ludwig Institute for Cancer Research. This work was supported by the NIH and by the Ludwig Institute for Cancer Research.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to I.M. (ira.mellman@yale.edu).

HIGHLIGHT ON THE NETHERLANDS



This section gives you an insight to a variety of scientific areas in the Netherlands including exciting and innovative centres of competence.

We will additionally introduce you to companies, institutions, research networks and universities offering world class and competitive recruitment opportunities at the very heart of Europe.

This feature is supported by the Bio Career Event

Bio
CareerEvent

naturejobs

nature publishing group **npg**

ADVERTISEMENT FEATURE



Innovation with R&D at its heart

In their efforts to drive the growth of their brands through fast, large-scale, exciting innovation, Unilever offers

a multitude of opportunities for highly trained Research & Development people. PhDs Gerda Feunekes and Matt Golding are among the dedicated professionals developing products designed to promote vitality and healthy living.

One could almost call Unilever, a brand behind brands. The chances are that this morning, you used Unilever soap in the shower, brushed your teeth with Unilever toothpaste, you dress in clothes washed with Unilever detergents, spread Unilever margarine on your toast and drank a cup of Unilever tea at breakfast and cleaned-up afterwards with a Unilever household cleaner.

If you look carefully on the packaging, you'll find the discreet 'U' of Unilever, backing up globally respected brands such as Dove, Signal, Omo, Becel, Bertolli, Knorr, Lipton, Ola and Cif, to name but a few.

R&D at the heart

These brands are subject to constant innovation, with R&D

at its heart, combining world-class science with deep consumer insight to produce new technology for consumers around the world – especially Unilever's new strides toward healthier products designed to help consumers *"look good, feel good, and get more out of life"*.

Based at the company's R&D facility in Vlaardingen near Rotterdam in The Netherlands, PhD Gerda Feunekes (PhD in Human Nutrition from Wageningen University, The Netherlands) and Matt Golding (PhD in Emulsion Stability from Leeds University, UK) have each spent almost ten years with the company, and currently work at the Unilever Food & Health Research Institute in Vlaardingen, serving Unilever's global food science.

Vitality

"I joined Unilever because I saw it as a great opportunity for contributing towards the best composition of our food products in line with the latest scientific developments," says Feunekes "When I joined, while we didn't have the core Vitality mission we have today, brands such as Becel, Flora, etc. were already focussing on that area. My PhD was about food choice, so it made me very suitable for our Consumer



ADVERTISEMENT FEATURE

Perception and Behaviour department. After five years, I moved to the nutrition area where I combined my skills in nutrition and my understanding of the consumer."

Applied techniques

"Having done a PhD in the area of emulsions, I had a particular interest in how food structures could be created, manipulated, and stabilised," says Golding. "I was lucky in the respect that my PhD was



industry supported, by Unilever. I spent a month during my studies at the Unilever research facility at Colworth House in the UK. The diversity of challenges we face within Unilever is remarkable. You are really being creative and coming up with new solutions," he says.

The Unilever R&D capability consists of six principal laboratories: two in the UK (Colworth House and Port Sunlight), one in the Netherlands (Vlaardingen), one in the US (Trumbull), one in China (Shanghai) and one in India (Bangalore). They work seamlessly with a network of global and regional technology centres that provide the next generation of products.

Healthy balance

Feunekes and Golding agree that one of the wonderful aspects of working for Unilever is that the company provides ample opportunity for personal development and training allowing you to grow within your role, develop yourself, and eventually grow out of your role, into new ones.

Unilever consistently ranks among the world's most admired employers and has a reputation for putting people first, providing opportunities for people to pursue their careers goals, develop professionally and maintain a healthy balance between their professional and personal lives.

For further information on careers within Unilever, visit www.unilever.com/rjobs or email to RecruitmentNL@unilever.com



W90520R

Amsterdam BioMed Cluster

Where do I go for biomedical business?

I amsterdam.

www.amsterdambiommed.com

Amsterdam, where Life Sciences and Commerce meet

Amsterdam BioMed Cluster is the network of universities, medical centres, specialized research institutes and service providers that fosters development of innovative drugs and technologies in the region. Amsterdam BioMed Cluster drives sustained growth of the biomedical industry, by stimulating entrepreneurship and promoting the flow of knowledge from academia to industry. It provides national and foreign companies with easy access to the Amsterdam life sciences and biotech community, and is managed from the offices of the Amsterdam Innovation Motor, which can be contacted at www.amsterdambiommed.com, or by phone 00 31 (0) 20 531 44 28.



Medical Business Park AMC and Science Park

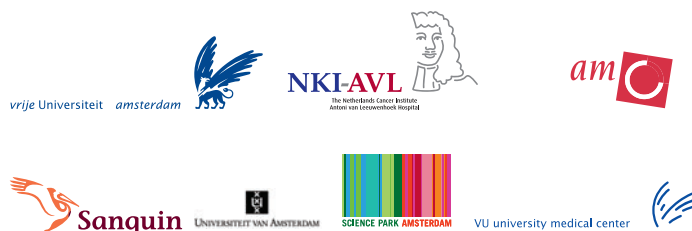
Amsterdam. The city of Amsterdam is developing two unique locations for the international life sciences industry: Medical Business Park AMC is located right next to the academic medical centre AMC and offers outstanding opportunities for enterprises active in (bio)pharmaceuticals, medical devices and patient care systems.

At Science Park Amsterdam an unique collaboration between scientific research, education and innovative businesses is reinforced by exploiting synergies between the life sciences and the computing communities. This park currently houses the Faculty of Science of the University of Amsterdam, several international renowned research institutes and over 80 high-tech enterprises.

Science Park Amsterdam and Medical Business Park AMC are currently offering actual office space in flexible multi tenant buildings, 20,000 m² new office and lab space will be realized within the next two years. In addition 250,000 m² building plots are available for innovative and high-tech businesses to realize their own offices and laboratories. Contact Ms. Lian Tjoa at ltjoa@amsterdamsience.com, phone +31-20-525 7873.

Scientific partners.

Years of tradition of excellence in research in oncology, neuroscience, autoimmunity, and cardiovascular and infectious diseases, have led to a dynamic environment for innovation and technology development. Companies will find easy access to technologies, clinical trial capabilities, patient cohorts and bio-banks, scientific and medical experts, and highly trained laboratory personnel.



Jobseekers and students.

Over 6000 students choose Amsterdam for Master and PhD courses in life sciences and medicine. The academic environment is international, with English as the classroom language. Scientists and technicians will find extensive job opportunities in the region in research institutes and one of the 80 biomedical companies. The BIOCareer Event on May 31, 2007 in Amsterdam will showcase opportunities in all of the Netherlands, see www.biocareerevent.nl.

Technology Transfer.

The Amsterdam BioMed Cluster is involved in the regional program to actively support new technological business initiatives, especially in life sciences. This program, co-financed by the government, and in collaboration with local business experts, stimulates universities and academic centers to commercialize their knowledge and inventions through spin-off companies. The program offers 'Techno starters' an attractive loan, professional coaches, access to research facilities and a network to optimally realize the first requirements to found a company and/or raising seed-capital. For more information please contact one of the technology transfer offices at liaison@uva.nl, info@tto.vu.nl or bkt@amc.nl.

For all inquiries: www.amsterdambiommed.com or + 31 (0)20 531 44 28

AIM Amsterdam Innovation Motor



Tipharma.com

Top Institute Pharma Drug Development Going Dutch

Pharmaceutical industry, the Dutch government and academia are joining up for a major pharmaceutical collaboration aiming to speed up the development of urgently needed medicines. The new organization, TI Pharma, is on the lookout for hundreds of entrepreneurial, ambitious researchers.

TI Pharma, a major public-private research consortium, officially launched in July 2007, is currently in the ramp-up phase. Projects are about to start and TI Pharma is eager to recruit ambitious young scientists. Top-notch research projects will result in an accelerated access to new medicines and vaccines. The program aims to target major human diseases (as defined in Priority Medicines (WHO 2004)) such as auto-immune diseases, cardio-vascular diseases, cancer, infectious diseases and brain diseases.

Approximately thirty biomedical companies, ranging from small biotech start-ups to large pharmaceutical corporations, have joined the initiative. Dutch universities are gearing up to kick off new research projects. Internationally acclaimed research institutes will soon follow suit.

Extensive training

The comprehensive TI Pharma program extends far beyond organizing and funding research. The institute is highly committed to developing a generation of scientists ready to tackle all aspects of the development of medicines. TI Pharma will offer both PhD and post doctoral students extensive training in industry-oriented topics such as drug discovery, intellectual property, business skills and regulatory compliance.

Looking ahead

With €60 million to spend next year, TI Pharma will rapidly generate a formidable wave of PhD and postdoc positions. There are bigger things to come over the next few years when EC funding is expected to expand the research coffers.

"TI Pharma promises to provide a unique model for encouraging drug development," says Scientific Director Daan Crommelin (Professor of Pharmaceutics, Utrecht Institute of Pharmaceutical Sciences, Utrecht University, Netherlands and adjunct professor at the Department of Pharmaceutics and Pharmaceutical Chemistry, University Utah, U.S.A.).

"So yes, we definitely have both European and international ambitions".

Coming up:

300+ research jobs

Through TI Pharma, about two hundred PhD and one hundred postdoc positions will soon open up at many prime Dutch research labs, offering young scientists the opportunity to extend their skills with industry-backed training. "Having demonstrable knowledge about drug regulation, intellectual property and business development will no doubt add a great deal to their CVs," says Daan Crommelin, TI Pharma's scientific director. TI Pharma especially welcomes postgraduate and postdoctoral students who are eager to turn cutting-edge basic research into real-world medicines. Because of the program's international orientation, they will need to have good English language skills.

Job opening alerts

While final details of most positions are still being worked out, those interested can send open applications with CVs to Karin Huiberts, human resources manager at TI Pharma. Karin also accepts subscriptions to e-mail alerts that will be issued when new job opportunities open up. Karin can be reached at huiberts@tipharma.com or +31-71-527-1430.

TI Pharma's Research Priorities

Therapeutic areas, guided by criteria set in the 2004 WHO report *Priority medicines for Europe and the world*.

- (Auto-) Immune diseases
- Cardio-vascular diseases
- Cancer
- Infectious diseases
- Brain diseases

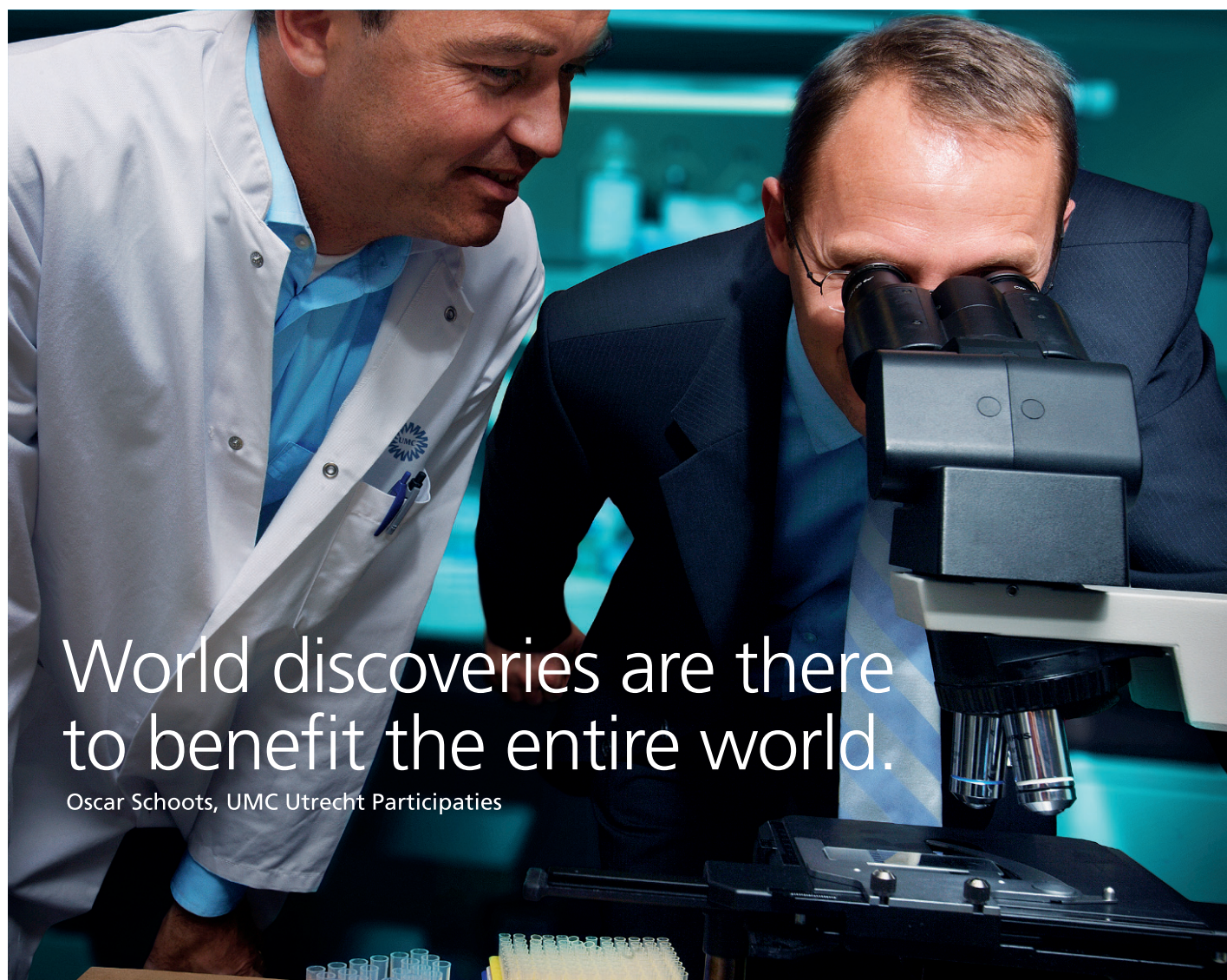
Enabling technologies:

- Therapeutic target finding, validation & animal models
- Lead selection and in-silico modelling
- Predictive drug disposition and toxicology
- Biomarkers and bio-sensing
- Drug formulation, delivery and targeting
- Pharmaceutical production technologies



W90092R

tipharma.com



World discoveries are there to benefit the entire world.

Oscar Schoots, UMC Utrecht Participaties

Working for UMC Utrecht means engaging in innovation and improvement no matter what the area or the job – from healthcare to research and education. One example is our business start-up support for scientists. For further information including vacancies, personal jobmail and work news film clips, visit werkenbijUMC utrecht.nl.



Scientists getting down to business. Making a scientific discovery is fantastic; but then what? Oscar Schoots: 'Publication, lecturing...what have you. Or you could start your own business. And since that's probably the one thing scientists don't know much about, we give them all the business support they need.'



University Medical Center
Utrecht

Always working to improve

W88647R

naturejobs

For ongoing recruitment communication needs, please contact your Netherlands representative

Reya Silao on E: r.silao@nature.com

T: + 44 (0) 20 7843 4970

Voor nadere inlichtingen op het gebied van communicatie en werving kunt u contact opnemen met uw vertegenwoordiger

Reya Silao, E: r.silao@nature.com

T: + 44 (0) 20 7843 4970



Organon is a global leader in the creation of innovative prescription medicines for fertility, gynecology, anesthesia and neuroscience, which contribute to the health of people and their quality of life.



Welcome to our world

Organon, the Akzo Nobel human healthcare business, with shared head offices in Roseland, NJ, USA and Oss, The Netherlands, is the largest innovative biopharmaceutical company of Dutch origin. The company's products are currently sold in over 100 countries. Organon employs around 14,000 people spread across more than 60 countries. Organon strives to remain or become one of the leading biopharmaceutical companies in each of its core therapeutic fields: fertility, gynecology, anesthesia and neuroscience. Research areas also include immunology and oncology.

Searching for new or better medicines

Ever since Organon was founded in 1923, it has had a strong scientific orientation. The company does large-scale research into innovative medicines, which has resulted in the development of a variety of products over the years. Organon is highly committed to its global R&D activities (between 18-19% of sales income is invested in R&D) and welcomes strategic alliances in order to help bring new drugs to the market. The Organon



strategy in biotechnology is to develop and market new biological entities within areas of interest such as immunology and oncology.

Working in a dynamic scientific environment

Organon R&D strives to create a stimulating and dynamic scientific environment for its workers. It invests in new technologies and collaborates with specialist technology companies to ensure that R&D programs are at the forefront of progress in science and medicine. R&D staff has state-of-the-art equipment and technology at their disposal, and is challenged to use their expertise and creativity to the fullest as members of interdisciplinary teams.

We have R&D facilities at 5 different locations in The Netherlands, USA, France, Scotland, Germany and Japan. Around half of our almost 2500 R&D employees work in The Netherlands.

Discover Organon by visiting: www.organon.com

3TU. Federation

Join our best researchers in pushing the frontiers of technology



The universities of Delft, Eindhoven and Twente are among the top in the world in several fields of technological research.

To meet the challenges of an evermore-competitive international research and funding environment, the three universities are joining forces. Supported by substantial funding, 75 of their highest ranked full professors will combine their research efforts in 5 Centres of Excellence.

Over the coming months, they will be looking for 28 new colleagues. Full professors who can help them push the frontiers of technology.

The 5 Centres of Excellence are:

3TU.Centre for Intelligent Mechatronic Systems

Microsystems - robotics precision technology - embedded systems
Coordination by Paul van den Hof, Henk Nijmeijer, and Stefano Stramigioli

3TU.Centre for Sustainable Energy Technologies

Solar energy - hydrogen production - fuel cells - bio refinery - storage

Coordination by Hans Kuipers, Tim van der Hagen, and Hans Niemantsverdriet

3TU.Centre for Dependable ICT Systems

Dependable computer networks - communication, information and security systems - ambient intelligence

Coordination by Peter Apers, Kees van Hee, and Patrick Dewilde

3TU.Centre for Multiscale Phenomena

Turbulence dynamics - multiphase flows - transport - integrated numerical-experimental approaches

Coordination by Dick van Campen, Detlef Lohse, and Gijs Ooms

3TU.Centre for Bio-Nano Applications

Single molecule and single cell studies - nano-sensing and bio-sensing

Coordination by Albert van den Berg, Bert Koopmans, and Huub Salemink

More on the new positions for professors and our ambitions can be found at www.3tu.nl

WCFS

TOP INSTITUTE FOR FOOD AND NUTRITION RESEARCH

The food industry is complex and increasingly driven to add value, whether through innovative ingredients, active packaging or safety, so there is a need to pool knowledge and resources. WCFS, as one of the four leading technology institutes in the Netherlands established in 1997, aims to be the networker and facilitator of breakthrough research in food and nutrition.

WAGENINGEN
CENTRE
FOR FOOD
SCIENCES



WCFS

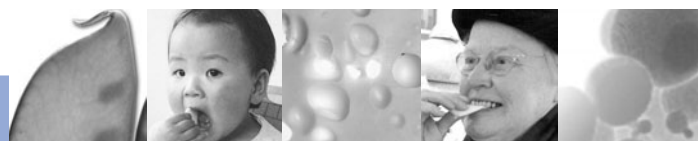
ALLIANCE OF GOVERNMENT, INDUSTRY AND RESEARCH

INDUSTRY

AVEBE | DSM | COSUN | NZO | CSM | UNILEVER

RESEARCH INSTITUTES

NIZO FOOD RESEARCH | TNO QUALITY OF LIFE | UNIVERSITY MAASTRICHT | WAGENINGEN UNIVERSITY AND RESEARCH CENTRE



The next phase in the 'Food and Nutrition Delta', a joint initiative between the food industry, research organisations and the government, aims to make the Netherlands one of the major suppliers of innovation for top quality food.

WCFS wants to be the European engine for this innovation. 'Within this framework our ambition to grow, substantially expand our research portfolio and attract partners from Europe's food industry is a realistic one,' says managing director Driek Vergouwen. 'It is based on the past pivotal role of WCFS in successfully enabling the food industry develop products and technologies that respond to consumer demand for safe, healthy and good-tasting food.'

Essential breakthroughs

Scientific, technological and commercial opportunities abound in the food industry. This is the context in which the top institute for food and nutrition facilitates research that reflects the major trends and issues affecting both the consumer and the industry.

This research involves key concerns in western society like obesity & metabolism, satiety, gastrointestinal health and mental performance. 'For example, we carry out research to develop products that contain specific proteins to make you feel full but yet with sufficient fat to make the food taste good,' explains programme director Professor Robert Jan Brummer.

WCFS also aims to provide the scientific basis for food technologists to design and control sensory and structural properties, or to engineer food to control food intake by inducing satiation and increasing the duration of feeling satisfied. 'You need your brain to produce *satisfied* feelings,'

says Professor Rob Hamer, programme director,' and this may be one of the reasons why people don't lose weight on *light* products. They don't feel satisfied, so they don't stop eating.'

Another important focus lies on providing new, breakthrough routes using genomic approaches and techniques. Programme director Professor Willem de Vos: 'In many cases the molecular structures of the bioactive ingredients are known, but we know very little about how these bioactive ingredients interact with other food components or what happens when they are eaten or what they actually do in the body.'

TANGIBLE RESULTS

There is tangible evidence of WCFS' success in all its research areas with the filing of more than 30 patents and a large number of published scientific papers. To the satisfaction of its partners. Director of life science products at DSM, Joop Roels:

'At WCFS we get access to high quality pre-competitive research that supports our own efforts in innovation in food and nutrition. This is vital as innovation in food and nutrition is one of the cornerstones of our mid and long term strategy. Furthermore, we also value the opportunity to keep in close contact with other industry and knowledge institute participants in WCFS. The institute provides a valuable opportunity for networking and is central to our policy of open innovation.'

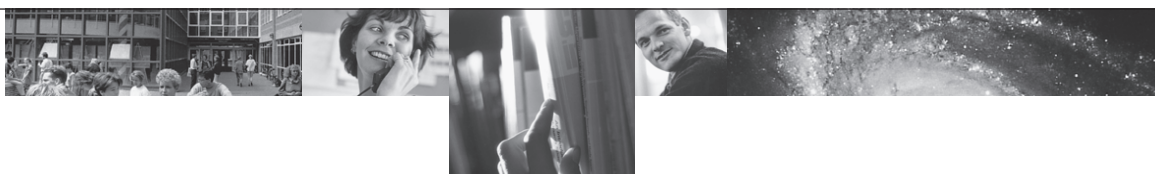


W69873R

Are you at the top in food and nutrition research?

Then we need you to come work with us!

Information about current vacancies can be obtained from our website www.wcfs.nl



Working at the frontiers of knowledge

Working at the University of Groningen (RUG) means working at the frontiers of knowledge. The RUG offers researchers and students the opportunity to expand those frontiers, to develop their talents and together with others realise top quality achievements. The RUG therefore deliberately opts for an interdisciplinary approach to knowledge. Discoveries are born and innovation realised at the interfaces between the various scientific fields. Researchers and students in Groningen benefit from the rich assortment of disciplines that the RUG has built up since its founding nearly four hundred years ago.

Knowledge is universal and transcends boundaries. The RUG has deliberately chosen to make a priority of intensive and worldwide cooperation with other leading universities and organisations. As well as being outward looking, the RUG is also closely involved with its own region where it is one of the largest employers. The university's 22,500 students and 6,000 staff have a distinct academic identity, firmly rooted in the wider social context. They are a vital element of the vibrant city of Groningen - a great place for students and staff.



RUG

The closing date for applications is **november 22, 2006**. Full applications, full CV plus a brief outline of current and future research and the names and contact details of two academic referees should be sent to: The University of Groningen, Personnel & Organisation Department, P.O. Box 72, 9700 AB Groningen, The Netherlands. Please state the vacancy number on the envelope and at the top of your letter.

Additional information about vacancies at the RUG is available on the university web site the university website:

(www.rug.nl)

The RUG has a special careers advisory service for the partners of new staff who move to the area.

**University Medical Center Groningen,
The Netherlands**

• Associate (or assistant) Professor in Experimental Nephrology

VACANCY NUMBER 206221

Job description:

We are seeking an outstanding scientist, with a solid experimental background (Biology, Medical Biology, or comparable expertise) as a structural staff member of the taskforce Nephrology of the UMCG. The successful applicant will join a multidisciplinary and successful scientific environment, that focuses on mechanisms and prevention of progressive renal function loss in native and transplanted kidneys. There is strong emphasis on translational research, and on training of junior investigators. The appointment coincides with establishment of a new laboratory unit for experimental nephrology that will be headed by the applicant.

Applicants should possess a PhD, and have several years of independent (or post-doc) research experience, and proven fund-raising capacities. Applicant will need to be in post by summer 2007. The appointment will in principle be as an associate professor, but candidates with the potential to qualify as an associate professor within five years, may apply for the position of assistant professor.

Informal inquiries should be made to the Chair of Experimental Nephrology, Prof. Gerjan Navis, e-mail g.j.navis@int.umcg.nl, phone +31 50 3614441 or +31 50 3612955.



Rijksuniversiteit Groningen

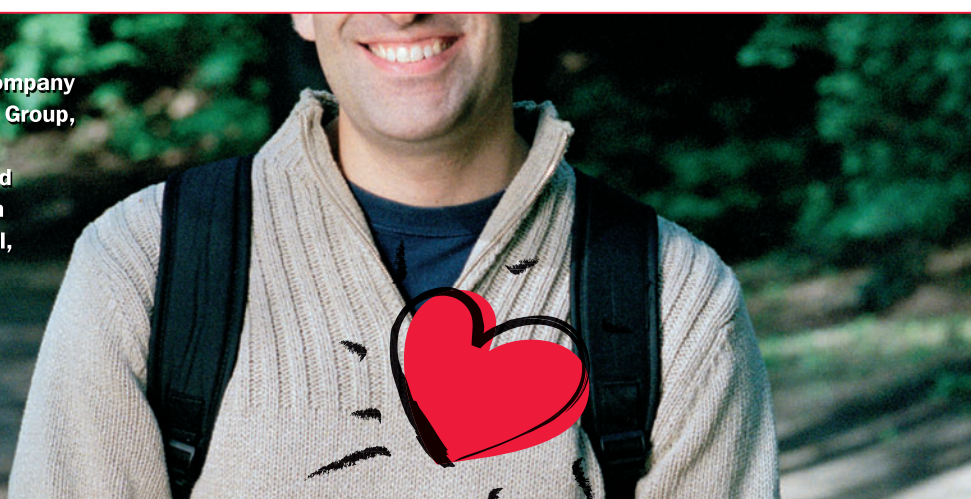
W90643R

**Bio
CareerEvent**

**Don't miss the Bio Career
Event in Amsterdam on the
31st May 2007!**

**For more information,
please go to
www.biocareerevent.nl**

"If I manage a project or company that is part of the Sara Lee Group, I run it as if it was my own business. I care for profit and I care for well-being. How am I going to achieve this? Well, to be honest, that's my business to."



Sara Lee is a versatile international company that produces and markets high quality consumer products and has a unique company philosophy: everyone has the freedom to carry out their own responsibilities as if it is for their own business. This requires passion for your work and your products. For the subsidiary Household & Body Care Research in The Hague/The Netherlands, we are looking for the following specialist:

Project Manager

As Project Manager you will take the lead in developing new products for the Bath and Shower product category to be marketed under international brands such as Radox, Sanex, Duschdas and Monsavon. You will have contact with external organisations and will work in dedicated teams with colleagues from Marketing and Operations. As well as an academic background in pharmacy, food science or chemistry you will be well experienced in Product Development and Project Management and have the capability to fulfil leadership function.

Look for more information on the position and how to apply on
www.workatsaralee.nl

We do not appreciate contact from employment agencies.



It's my business

W90568R

Radboud University Nijmegen Medical Centre is a leading academic centre for medical science and health care. Knowledge forms the heart of our organisation, connecting research, education and patient care. Over 8,000 staff and 2,500 students are committed and ambitious to contribute to the future of medical science, education and health care.

Radboud University Nijmegen  **Medical Centre**



Marlies Welling, surgical technician student

"Never a dull moment in this hospital. Students get the opportunity to learn a lot. Colleagues are ready to help you any time. One feels that this is a genuine teaching clinic."

The departments of Cell Physiology, Molecular Hemato-Oncology and Antropogenetics invite

Postdoctoral researchers and PhD students

with a strong background in chemistry, molecular, medical or cell biology, to apply on different projects:

- **molecular mechanisms and origin of (kidney)cancer**
- **the mechanisms and oncogenesis of leukaemia**

The departments perform research on molecular and cellular physiology of water and electrolyte transport, molecular mechanisms that regulate leukaemic stem cell proliferation and differentiation and contributive factors to the development of acute myeloid leukaemia.

Also visit our Job Opportunity Market on February 13th 2007

W90639R

Please download further details of application procedures at

www.umcn.nl



Leiden/Amsterdam Center for Drug Research
 P.O. Box 9502, 2300 RA Leiden, The Netherlands
 Tel +31 71 527 4341, Fax +31 71 527 4277
info@lacdr.leidenuniv.nl

TI Pharma is a major Dutch public-private initiative, accelerating development of WHO-designated 'priority medicines' through high-quality translational research.

The **Leiden-Amsterdam Center for Drug Research (LACDR)** is one of TI Pharma's key players. Combining strengths of Leiden University and VU Amsterdam, LACDR offers highly competitive research and education in drug discovery and drug development.

LACDR focuses on drugs targeting cancer, (auto) immune diseases, cardiovascular diseases and brain diseases. **Research programmes** include Molecular & systems pharmacology, Medicinal chemistry, Bio-analytical sciences, Drug delivery & targeting, Molecular toxicology & drug safety, and Advanced PK-PD modelling.

For upcoming research projects, LACDR is looking for
**Postdoctoral Research Fellows
 &
 PhD candidates**

Individual vacancies and contact details will be published on www.lacdr.nl. Prospective candidates can also subscribe to TI Pharma's general job alert mail service at huiberts@tipharma.com.



tipharma.com

W90553R

Amsterdam - Helsinki - Barcelona - Braga - Edinburgh



Series of Conferences and Training Courses

1ST CONFERENCE - MOLECULAR PROFILING OF THE GENOME
 VUMC Amsterdam NL, May 2nd - 5th, 2007

Topics:

Epigenomics
 In silico genome analysis
 Molecular genetics of disease*
 Genome profiling to identify disease* genes
 Structural variation of the genome
 Post-genomic technologies

* Infectious diseases, cancer, mental retardation, immunology

Speakers:

James Lupski (Baylor College of Medicine, TX, USA)
Donna Albertson (UCSF, CA, USA)
Anne Kallioniemi (University of Tampere, Finland)
Elisabeth Blennow (Karolinska University, Sweden)
Erwin Schur (McGill University, Canada)
Matthew Hurles (Wellcome Trust Sanger Institute, UK)

Complete reimbursement for eligible researchers

W90674E



Get the latest jobs while
 playing with your BlackBerry®.

BlackBerry is a trademark of Research In Motion Limited.

naturejobs

nature publishing group **npg**



SPECIAL FEATURE

THE NATIONAL INSTITUTES OF HEALTH WWW.NIH.GOV

An Institute on the Move

A Conversation with the Director of the NIEHS

Just over a year ago, David Schwartz took over as director of the National Institute of Environmental Health Sciences (NIEHS) and the National Toxicology Program. In his first year, he instituted a process to develop a Strategic Plan* to guide the institute's research programs over the next five years. This plan is now being put into action.

You are the first new director for the NIEHS in almost 15 years. What perspective do you bring to this position that's new and different?

I think it's the perspective of a physician-scientist. Actually, the first two NIEHS directors were also MDs, but it's been a long time since the institute was led by someone with a background in clinical and translational research. In environmental sciences, there is a huge need to bridge the gap between basic sciences and human health. As a pulmonologist, I've seen firsthand the continuum from toxic exposure to diseases such as asthma and other chronic airway diseases. From my perspective, I think the most exciting, and more importantly the most applicable, discoveries for health are going to take place in the middle of this pathway, in the area of translational research.

Where is the NIEHS headed as an institute?

We're embarking on a new paradigm for research in environmental health that focuses less on the toxic chemical or agent and more on the result of that exposure in human beings. Because of the fantastic work that's been done at this institute over the last 30 years, we know a lot about environmental agents. What we don't understand well in most cases is exactly how they work with genes in human systems to cause disease. By focusing on what's happening in people and working backward from there, if you will, I think we can really hope to get to a new level of understanding of disease. And from there to developing ways to markedly improve people's health.

Some may see this as a radical change in trajectory for the institute. Is it?

I think rather than a change in direction, it's really a natural extension of the previous work. The more we know in science, the more opportunities and challenges there are. This is particularly true at this point in time. We finally have the tools and technology to understand how environmental exposures are affecting human biology and altering the risk of disease, especially in complex diseases like asthma and autism that are caused by multiple environmental and genetic factors. We're starting from the mountain of knowledge and data we have, and are continuing to acquire, and

following a plan to drill deeper into the problems of disease than ever before. It's an approach that takes the best basic research on what people are exposed to and how they are exposed, combines it with the clinical understanding of disease phenotypes and pathogenesis, and focuses them together on figuring out where the nexus is, that place where complex disease occurs. In a sense, it is a radical idea, since these two scientific camps haven't really come together in this way before.



Steve McCaw/Image Associates

How exactly will the Strategic Plan be accomplished and when can we expect to see results?

The critical components for success are outlined in the plan, and I encourage people to read it, but the core of our efforts will focus on building integrated environmental health research programs that are truly interdisciplinary. This gets to the creative thinking and cross-pollination of ideas that leads to discovery. To that you have to add development of and access to the most advanced scientific technology and infrastructure, as well as training of young scientists. And finally, you have to make sure that you get out of the Ivory Tower and are constantly checking in with what's happening in the real-world environment so that your assumptions remain accurate and relevant, and you remember the end goal, which is to alleviate suffering and improve health. Although this was conceived as a five-year plan, scientific discoveries don't necessarily adhere to anyone's timeline. But we know the sooner we start down the path the better.

What's the one most critical element necessary for the NIEHS to be successful?

I would say the willingness of the scientific community to expand our notion of how we do what we do. To be open to new ways of thinking about the problems and new ways of approaching how we solve them. And this involves being willing, and in fact eager, to bring our skills, talents, and ideas to the table with our colleagues. If we can do that, I really believe that we'll have an enormous impact on the future of human health.

*<http://www.niehs.nih.gov/external/plan2006/home.htm>



Tenure-Track Principal Investigator, Laboratory of Biochemistry and Molecular Biology Center for Cancer Research, NCI

The Center for Cancer Research (CCR), National Cancer Institute (NCI), National Institutes of Health (NIH), Department of Health and Human Services (DHHS) is accepting applications for a Tenure-Track Investigator in the Laboratory of Biochemistry and Molecular Biology (LBMB). The CCR's mission is to reduce the burden of cancer through exploration, discovery, and translation. As part of this mission, the LBMB investigates basic cellular processes, with an emphasis on the biology and biochemistry of chromosomes and the cell nucleus. LBMB fosters a highly interactive and collaborative environment that is supportive of junior investigators. Investigator research interests include:

- Carl Wu**—Chromatin remodeling, transcription, and functions of histone variants
- Yawen Bai**—mechanisms of protein folding and protein dynamics
- Dhruba Chatteraj**--maintenance of multiple chromosomes in *Vibrio cholerae*
- Shiv Grewal**—RNAi and higher-order chromatin assembly in *S. pombe*
- Michael Lichten**—meiotic recombination and chromosome structure in *S. cerevisiae*
- Bruce Paterson**—myogenesis and regulatory RNA pathways in *Drosophila*
- Yikang Rong**—DNA repair and telomere maintenance in *Drosophila*

Candidates should have a Ph.D. or M.D. degree with proven ability to conduct innovative research in the broad areas of genome expression, transmission, and stability. The successful candidate will perform independent research funded by the NIH Intramural Research Program as a member of the LBMB, and will also participate in the CCR Center of Excellence in Chromosome Biology (CECB). More information can be obtained at the following the websites:

LBMB--<http://ccr.cancer.gov/labs/lab.asp?labid=783>, CECB--<http://ccr.nci.nih.gov/initiatives/CECB/>.

Send C.V., statement of research plans, and 3 letters of recommendation by **December 4, 2006** to: **NCILBMBsearch@mail.nih.gov** or to **Zoraida Villadiego, NIH, Bldg 37 Rm 6106C, MSC 4260, Bethesda, MD 20892**. Electronic submissions are encouraged.



HELP US HELP MILLIONS

National Institutes of Health National Institute of Allergy and Infectious Diseases ImmunoTechnology Section, Vaccine Research Center

Postdoctoral Researcher

The Dale and Betty Bumpers Vaccine Research Center (VRC), NIAID and NIH bring together a diverse group of scientists with the goal of advancing vaccine-related research as well as initiating and advancing vaccine candidates through clinical trials. The ImmunoTechnology Section is a highly active and collaborative laboratory with several postdoctoral fellows, students, and biologists. For a detailed description of our laboratory's program, see selected publications (right) and http://www.vrc.nih.gov/VRC/labs_immunotechnology.htm.

Currently, we have a postdoctoral position available to study mucosal or central immune responses. We are using the nonhuman primate model (SIV infection) to define viral and cellular dynamics across all tissues and use state-of-the-art multicolor flow cytometers as well as cellular and molecular technologies uniquely available at the VRC and the NIH.

We are seeking a highly motivated and creative individual, with a Ph.D. or M.D. and experience in cellular or molecular immunology, to define and head a unique research program within the laboratory. Candidates with any level of training are encouraged to apply. Interested candidates should send a copy of their CV, and contact information for three references to:

Mario Roederer
VRC, NIAID, NIH
40 Convent Dr., Room 5509
Bethesda, MD 20892-3015
301-594-8491; FAX 301-480-2651
Email: Roederer@nih.gov

Selected references:

- Mattapallil, J. J. et al. Massive infection and loss of memory CD4 T cells in multiple tissues during acute SIV infection. *Nature* 434, 1093-1097 (2005).
- Mattapallil, J. J., et al. Vaccination preserves CD4 memory T cells during acute simian immunodeficiency virus challenge. *Journal of Experimental Medicine*, 203, 1533-41 (2006).
- Chattopadhyay, P. K., et al. Quantum dot semiconductor nanocrystals for immunophenotyping by polychromatic flow cytometry. *Nature Medicine*, 12, 972-7 (2006).
- Perfetto, S. P. et al. 17-Color Flow Cytometry: Unraveling the Immune System. *Nat Rev Immunol* 4, 648-655 (2004).
- Chattopadhyay, P. K. et al. A live-cell assay to detect antigen-specific CD4 T cells with diverse cytokine profiles. *Nature Medicine* 11, 1113-1117 (2005).
- Nishimura, Y. et al. Non-activated naive CD4 T cells are the principal source of progeny virus during acute infections of monkeys with highly pathogenic X4-tropic SHIVs. *Proc Natl Acad Sci (USA)* 102, 8000-8005 (2005).



Vaccines for Life™

VACCINE RESEARCH CENTER
National Institute of Allergy and Infectious Diseases
National Institutes of Health
Department of Health and Human Services





Tenure-Track Position National Institute on Deafness and Other Communication Disorders

The Division of Intramural Research, National Institute on Deafness and Other Communication Disorders (NIDCD), located in Bethesda, MD, is seeking a tenure-track scientist to establish an independent research program to study molecular and/or cellular mechanisms of hearing and balance. We welcome applications from candidates with a wide range of expertise. Preference will be given to candidates whose experimental approaches complement those of our existing strong programs in the genetics, development and cell biology of hearing. The successful candidate will join a dynamic group of scientists in a growing intramural program that is at the forefront of research on communication disorders.

The NIDCD offers an exceptional working environment including well-equipped research laboratories and numerous opportunities for collaboration. Candidates for this position must possess a Ph.D. and/or M.D., post-doctoral experience, and an outstanding publication record. Salary is commensurate with education and experience.

Please submit a curriculum vitae including bibliography, three reprints of recent relevant publications, statement of research interests, an outline of your proposed research, and the names and addresses of three references to: **Ms. Trudy Joiner, Office of the Scientific Director, NIDCD, 5 Research Court, Room 2B28, Rockville, MD 20850 (joinert@nidcd.nih.gov).** Applications will be accepted until **December 15, 2006**.



Staff Scientist Position Available

The National Institute of Allergy and Infectious Diseases, a major research component of the NIH and the Department of Health and Human Services, is recruiting a Staff Scientist. The position will be available in the Respiratory Viruses Section of the Laboratory of Infectious Diseases, and scientists with a M.D., D.V.M., or Ph.D. are eligible. The research activity involves (1) examination of the pathogenesis of pandemic and potential pandemic strains of influenza and their evaluation in vitro and in experimental animals; (2) influenza viral genomics, and examination of viral evolution in fitness and host adaptation; and (3) the development of influenza clinical trials in humans. This full-time research position offers a unique opportunity to work on investigations that range from basic molecular biology to clinical research. Staff Scientist applicants should have at least six years of laboratory work experience in molecular and classical virology research; the salary range is \$73,178 - \$165,195. Preference will be given to candidates who have experience working with avian influenza viruses and those with BSL3 experience. Applicants should submit their curriculum vitae, a letter of research interests, and names and addresses of three references to: **Jeffery K. Taubenberger, MD, PhD, Attn: D. Kyle, NIAID, NIH, Bldg 50 Room 6234, MSC 8007, 50 South Drive, Bethesda, MD 20892-8007, FAX: (301) 496-8312, email: dkyle@niaid.nih.gov**

Review of applicants will begin on **October 30, 2006** and continue until a successful candidate is identified.



Tenure-Track Principal Investigator Laboratory of Cellular Oncology Center for Cancer Research, NCI

The Laboratory of Cellular Oncology (LCO), Center for Cancer Research, of the National Institutes of Health, invites applications for a tenure track or tenure eligible principal investigator position in the area of research on papillomavirus biology, including vaccines. The LCO, which occupies recently renovated laboratory space, fosters an interactive research environment and the use of diverse experimental approaches and model systems. Applicants should have a Ph.D. and/or M.D. degree, a strong publication record, and demonstrated potential for imaginative research. Salary will be commensurate with education and experience. A one- or two-page statement of research interests and goals should be submitted in addition to three letters of recommendation and a curriculum vitae to: **Theresa Jones, Laboratory of Cellular Oncology, National Cancer Institute, NIH, Building 37, Room 4106, Convent Drive MSC 4263, Bethesda, MD 20892-4263; phone: 301-496-9513; fax: 301-480-5322; email: jonest@DC37A.nci.nih.gov.** Candidates must be U.S. citizens or permanent residents.

Applications must be received by **11/22/06**. The National Cancer Institute is an Equal Opportunity Employer. Selection for this position will be based solely on merit, with no discrimination for non-merit reasons such as race, color, religion, gender, national origin, politics, marital status, physical or mental disability, age, sexual orientations, or membership or non-membership in an employee organization.



Tenure-track Physician Clinical Center/Nuclear Medicine Department

This position is located in The Warren G. Magnuson Clinical Center, Nuclear Medicine Department (NMD).

We are seeking a research-oriented physician for a possible tenure-track position. An M.D. or M.D. /PhD with U.S. Nuclear Medicine Board certification and CT training is needed to provide diagnostic and therapeutic nuclear medicine procedures as well as to participate in clinical research protocols of the NIH Intramural Program. U.S. citizenship or permanent residency status is required.

Please submit your curriculum vitae, bibliography, and a letter describing your clinical, research, and management experience to: **Mrs. Veronica Olaaje, HR Specialist, DHHS, NIH, OD/CSD-E, 2115 E. Jefferson Street, Rm. 2B209 MSC-8503, Bethesda, MD 20892-8503. Phone: 301-435-4748. Email: volaje@mail.nih.gov.**

Salary is commensurate with experience. This appointment offers a full benefits package (including retirement, health, life and long term care insurance, Thrift Savings Plan participation, etc.). Application packages should be submitted as early as possible, but no later than **December 31, 2006**.

Selection for this position will be based solely on merit, without discrimination for non-merit reasons such as race, color, religion, sex, national origin, politics, marital status, sexual orientation, physical or mental handicap, age or membership or non-membership in an employee organization.

NW90387R

naturejobs

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

What happens when a country produces too many postdoctoral students? The answer, sadly, is that the students' career prospects suffer. When a particular field is flooded with postdocs, the students often end up trudging from lowly paid fellowship to lowly paid fellowship — with permanent positions proving elusive at best. Some students even countenance the unthinkable: abandoning science altogether, despite the years of education and work they have invested.

This is what was happening in the United States seven or eight years ago — especially in molecular biology. Ominously, it looks likely to occur in Japan very soon. Japan has some 10,000 postdoctoral fellows living either at home or in the United States, and does not seem to have sufficient capacity to give them secure employment, especially in academia.

Japanese postdocs should be able to head off impending problems by taking a lead from their US counterparts. US postdocs formed grassroots associations, which allowed them to pool their resources and learn about alternative career paths. As a result, many have become more flexible and mobile, both geographically and in terms of switching between sectors. So far, only eight such associations exist in Japan — although there is a chance that more may be on the way.

But Japanese students have to overcome some additional obstacles compared with US postdocs. They face a greater stigma if they leave academia for industrial research or off-the-bench careers, for example. And they may be more constrained by language limitations, as most international labs operate in English.

Nevertheless, this generation of Japanese postdocs has some advantages over those in the United States, many of whom were caught off-guard by sudden career restrictions. The Japanese postdocs can see the problem coming and prepare for it. The pioneering scientists who overcome the cultural and linguistic constraints of pursuing non-traditional science careers could well end up more secure than those who opt for the long, uncertain road to tenure-track positions.

Paul Smaglik, *Naturejobs* editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Assistant Editor: Gene Russo

US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager/Corporations:
Peter Bless
Classified Sales Representatives
Tel: +1 800 989 7718
East USA/Canada: Andrew Bennie
**NIH/Maryland/New York/
Pennsylvania:** Shelley Cohen

San Francisco Office
Classified Sales Representative:

Michaela Bjorkman
West USA/West Corp. Canada
225 Bush Street, Suite 1453
San Francisco, CA 94104
Tel: +1 415 781 3803
Fax: +1 415 781 3805
e-mail: m.bjorkman@naturesf.com

European Head Office, London
The Macmillan Building,
4 Crinan Street, London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
Advertising Production Manager:
Stephen Russell
To send materials use London
address above.
Tel: +44 (0) 20 7843 4816

Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
Naturejobs web development:
Tom Hancock
Naturejobs online production:
Catherine Alexander

European Satellite Office
Patrick Phelan
e-mail: p.phelan@nature.com

Japan Head Office, Tokyo
Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku,
Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com

MOVERS

Hans-Olov Adami, Chair, Department of Epidemiology, Harvard School of Public Health



2006–present: Professor of cancer epidemiology, Karolinska Institute

1997–2005: Professor and chairman, Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden

1990–97: Professor and then chairman, Department of Cancer Epidemiology, Uppsala University, Sweden

Hans-Olov Adami believes that today's students have a dangerous preoccupation with career planning. "Our mission isn't to publish papers, our mission is to make discoveries," he says. Adami has followed this credo throughout his own career. Although trained as a medical surgeon, Adami followed his passion to become one of the top medical and cancer epidemiologists in the world.

At Uppsala University in Sweden, Adami spent 17 years practising medicine and surgery. But it was his doctoral dissertation project — on the epidemiology of breast cancer — that sparked his interest in a new career path. Lacking mentors at Uppsala, Adami taught himself epidemiological study design. "All of a sudden, I realized how medical knowledge is generated," he says, emphasizing the often uncertain relationship between epidemiological research, which pinpoints the causes of disease, and clinical research, which diagnoses and treats. With clinical research colleagues, he began his first series of epidemiological and clinical projects soon after completing his PhD.

In 1986, he was appointed to a new senior research position in cancer epidemiology at Uppsala University, funded by the Swedish Cancer Society. He subsequently surrendered his surgical gloves to pursue research. The programme's success culminated in Adami's appointment as national chair of cancer epidemiology in 1990. Wanting more security for his employees, Adami made the bold decision to move the entire department to the Karolinska Institute in Stockholm. The department grew to more than 160 people, established the largest unit for biostatistics in Sweden and expanded into areas outside cancer, such as reproductive epidemiology.

Meanwhile, Adami forged connections with Harvard's school of public health as an adjunct professor. From next February, he will chair the department of epidemiology, and plans to use this opportunity to expand on global projects with non-Western countries such as Singapore. Current chair Meir Stampfer, whose work has entered controversial areas such as the link between obesity and mortality, says Adami's international perspective is unique and will help him lead the department in an increasingly global field.

Adami has already created a new training module that will divide trainees' time between Sweden and Singapore. He hopes it will help young people embrace their curiosity, rather than their perceived career goals. ■

Virginia Gewin

BRICKS & MORTAR

A bench to call your own

Aspiring entrepreneurs at the University of California celebrate the first birthday of an unusual biotech incubator this month. The QB3 Garage lets them rent the equivalent of a single lab bench, right in the heart of a premier research institute.

Incubators are usually located near university labs, but rarely in the same building. Yet in a five-storey building at the University of California, San Francisco (UCSF), six tiny companies split a 232-square-metre space, which includes a common laboratory plus private sections for offices or labs.

Named after the garage in Palo Alto where Bill Hewlett and David Packard launched their business in 1939, the QB3 Garage is the brainchild of the California Institute for Quantitative Biomedical Research (QB3), which has sites at UCSF and at the Berkeley and Santa Cruz campuses. The state government funds QB3 to help academics commercialize research.

The Garage helps companies through the transition from research idea to new business entity by keeping them close to potential academic collaborators and by leasing very small spaces, so firms don't pay rent for space they don't need. Also, the Garage doesn't take equity or intellectual property from its tenants.

The two founders of MynSys

Cellular Devices, which is developing a microknife for cellular surgery, rent the Garage's smallest space, 12 square metres. "We could probably do everything we're doing now at a different place, but it would cost us more, and not just in terms of money," says ophthalmologist David Sretavan. He adds that the location aids academic collaborations and the intellectual environment fosters innovation.

Tenants must be affiliated with research at QB3 or the university, their work must have biomedical applications, and QB3 spin-offs have priority as space is limited. Most tenants are funded by small business grants from the US government.

Leases last a mere 18 months. The goal is to give nascent companies a chance to perform proof-of-concept experiments or create prototypes before finding additional investment and moving on, explains Douglas Crawford, associate director of QB3. It's worked so far for the first tenant, Fluxion Biosciences, which got seed financing in September and plans to move to larger quarters in 2007.

Meanwhile, Crawford says he's had to turn down five applicants in as many weeks for lack of space, and other institutions have contacted him about starting garages on their campuses. ■

Monya Baker

GRADUATE JOURNAL

The joys of communication

Throughout my university years, I've spent a lot of time teaching and writing popular-science articles for magazines and newspapers. Sometimes people wonder why I bother. It's true that without such activities I might have graduated sooner. But I think scientists have an obligation to educate the public. Even in a highly educated country such as Finland, up to a third of the population does not believe in evolution, according to a recent European Union survey.

Apart from that, there are perks to communicating science. It allows me to stay in touch with fields of biology unrelated to my research. I've had the chance to observe weather conditions in space, learn how sociologists analyse political jargon and visit the inside of a magnetic resonance brain scanner. I am inspired by these reminders of how exciting scientific endeavours can be.

Teaching improves my understanding of science because it forces me to think things through. Sometimes I discover a supposedly basic idea that is not clear to me at all. When I'm lost in the fog of my own research, a student may ask a question that actually helps me to see more clearly. And best of all, teaching is a great way to find the connection between concepts. As I write the introduction to my thesis, trying to link abstract modelling, population genetics and behaviour, I'm grateful for this skill. ■

Katja Bargum is a graduate student in the Department of Biological and Environmental Sciences at the University of Helsinki, Finland.

A concrete example

When the cracks begin to show.

The Artistic Forms and Complexity Group: J.-P. Boon, J. Casti, C. Djerassi, J. Johnson, A. Lovett, T. Norretranders, V. Patera, C. Sommerer, R. Taylor and S. Thurner

"You paid what? I have a chunk of the Berlin Wall at home that's at least as 'artistic' as this," sputtered Bill. "Did you leave your brain behind when you abandoned physics for finance? Putting a paving stone on a pedestal and calling it a work of art doesn't make it so, Dave. It's like writing down the alphabet and calling it the 'theory of everything.'"

"I stare at financial data all day long, watching the markets rise and fall," Dave replied. "I get paid to spot patterns in all of that craziness. I know the mathematics behind its complexity. But stocks aren't beautiful — not like that crack in the rock. Its complexity catches the eye. You're the art dealer, Bill; surely you can spot natural beauty." Bill shrugged: "I never said art and complexity were the same thing."

"Nanosculpture!" The gallery director had said. "Fill the biggest space in London with stuff so small you can't see it? Protein ready-mades! Very funny! I predict a dismal flop!"

Bill wasn't sure if they hadn't gone too far this time. Proposing an environmental installation on the beauty of nature might have been better. But didn't Dave pay £100,000 just for a piece of cement sculpture? An invisible nanosculpture for Piccadilly Circus! You can't beat the complexity of that! Isn't art about pushing the boundaries? To hell with beauty and logic!

The mood at the board meeting was explosive. Most members thought nanosculpture was a bad joke. But the director made clear that the goal of this seemingly absurd project was to trigger a mix of thoughts, emotions — even aggression — in the minds of millions of people. The invisible (and extremely expensive) nanosculpture would 'materialize' much more than any conventional work of art. With these dynamics, the piece could be sold as a complex system. Unfortunately, the director could not convince Dave, a complexity specialist, to verbalize this at the meeting. Following a day of sometimes hostile debate, no resolution was reached.

Dave left the meeting, wondering how to formalize his sensations. Beauty comes first, something I can see through a form. Why not say the complexity of this invisible sensation is virtually a nanosculpture,

complex but invisible? Wouldn't this trigger mental processes equally well as the invisible nanosculpture? The most incredible phenomena in nature are invisible, but are mysterious works continually in progress. What about art? Is an invisible provocation enough? Isn't this the space of thought?

Bill's phone rang at an ungodly hour. He heard Dave mumble: "Janus."

"You mean Dionysus and Apollo?" Bill probed.

"No, art and science."

"Two faces looking in opposite directions?"

"But connected to the same device, a complex machine, actually."

"You mean art and science share the same...?"

"Yes, the same complexity," Dave said, abruptly hanging up.

Bill was at Heathrow, heading for New York. At the meeting in London, Dave crumbled, then whispered about some mystical nano-experience. Bill tried to contain his anxiety: could Dave work out his reaction to the nanosculpture project? The director was relying on him to persuade Dave to arbitrate between the factions on the board. But Dave's mind seemed to be splintering like that rock he'd installed in his living-room.

At the door, Bill found himself facing a woman in a white dress, no make-up, his own startled face reflected in mirrored sunglasses of the type he hated.

"I'm Bill," he stammered.

She cut him off: "Dave wants you to wait in this room," indicating a closed white door. "Make yourself comfortable," she said, and left him in a windowless all-white room: walls, ceiling, floor and a single piece of furniture: a white swivel stool. Even though the room was soundless, Bill soon felt a music-like tinnitus ringing in his ears, something he'd never experienced before. But the most remarkable disturbance was visual. Everything around him was white. But the surfaces weren't smooth. The impression was a dynamic irregularity that made him wary. Rising from the stool, he slowly approached

the facing wall, realizing that not even the faintest outline of a door was discernible. He must have dozed off, because suddenly he found himself looking up at Dave's face.

"Bill, your hour is up," he smiled.

"I've been here for an hour? I can't believe it."

"A full hour," said Dave. "Nano-art does that to you. Why don't you explain that to the director?"

Ultimately, Bill's power of persuasion had been so strong that he single-handedly brought the nano-installation into existence. But that was a year ago.

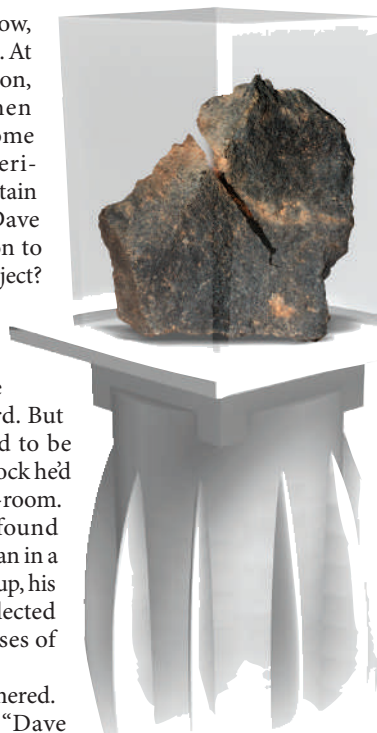
"It was entirely of your making," announced the director. "Now there's a bill to be paid. They've sued us for £478,356,298 for nano-vandalism." She went on to explain that he had unwittingly added a twist to the

nanosculpture: the tiny self-reproducing protein had been engineered to mimic prion-like behaviour. Each time someone walked across Piccadilly Circus while thinking about the unbearable complexities of life, a protein molecule would make a little crack in the stones in front of the person having the thought. Within weeks the entire pavement and many houses in the neighbourhood had cracked open completely. A spectacular work of art was now called *Mad Cement Disease*. "You will have to pay," the director explained. "But the suit does include VAT on the reconstruction costs.

And you get to keep the stones.

Perhaps Dave will buy them from you." ■

This story is condensed from one created in the Japanese 'Renga' fashion by ten writers, each contributing one paragraph in sequence, with only the coordinator knowing the source of each paragraph. The authors were at a workshop on "Artistic Forms and Complexity", partially supported by the European Commission and organized by John Casti at the Wissenschaftszentrum Wien in Vienna in October 2005. The original story is at www.wzw.at/file_upload/Kovacic_tmpphpcqCq5Y.doc.



JACEY

Abstractions



PROJECT LEADER

Light falling on a tiny mirror exerts a pressure, causing the mirror to move and potentially revealing its quantum-mechanical behaviour.

But at room temperature the effects are masked because the mirror is in a constant state of thermal agitation. Markus Aspelmeyer at the Institute for Quantum Optics and Quantum Information in Vienna, Austria, and his team have found a way to cool the micromirror using radiation pressure, and thus damp down this random agitation. Furthermore, their system can self-regulate its own cooling, removing the need for any active feedback loops. Their work on page 67 creates an important avenue for physicists to combine micromechanical motion and radiation to improve the sensitivity of high-precision measures — or one day even demonstrate quantum behaviour.

Why has no one achieved self-cooling by radiation pressure in a micromechanical system until now?

There have been attempts over the past 5 to 10 years, but no one had a micromirror that was lightweight enough, or highly reflecting. We were lucky to collaborate with guys from the United States and Austria who could work with our initially rough ideas for such high-sensitivity systems — which are only now possible because of nano- and microfabrication techniques.

What was the key breakthrough?

It was really a series of minor breakthroughs, but we did celebrate at our first successful attempt to fabricate the mirror used in the experiment. Our celebratory, now empty, bottle of champagne has a plot of our first characterization of the mirror on it.

The paper was put on hold while control experiments were reworked — during the World Cup no less. What happened?

We were ambitiously trying to provide the first full quantum description, rather than follow previous classical-physics attempts, so we had to be extremely careful. At one point we realized there seemed to be a factor of two missing when we compared our findings with results in the classical literature. It turned out that our original data were correct, and luckily we were able to settle all experimental and theoretical issues before the World Cup semi-finals.

What's the next step?

Classical physics offers no means to describe the quantum ground state, or lowest possible energy state. By continuing the work described in this paper and eliminating the current mirror imperfections, we will get there straight away. ■

MAKING THE PAPER

Eberhard Fetz

Computer-chip implants could help people with brain damage.

As part of a neuronal relay race, electrical signals travel from the brain to the muscles that control movement. Break any neuronal connection in the chain — through injury or disease — and the control may be lost.

Eberhard Fetz and his colleagues at the University of Washington, Seattle, attempted to use a small computer chip to replace these lost connections. But in experiments described on page 56, they found that their device could also serve to strengthen and rewire existing ones.

The collaboration between Fetz, Andrew Jackson and Jaideep Mavoori began in 2003, when Mavoori, an electrical engineer, was developing a computer chip that could be implanted in moths to record and stimulate neuronal activities as they flew.

"We heard about the work and thought something that small should be easy to implant into a monkey's brain," recalls Fetz. That insight turned out to be correct, although implementing it was not easy.

The team spent more than a year redesigning the chip so that it could be programmed to record signals from neurons firing in an area of the brain's motor cortex responsible for moving certain wrist muscles in one direction (site A). The chip, which was positioned on top of the brain, then delivered these signals a few milliseconds later to a nearby site responsible for controlling a different set of wrist muscles (site B). The chip made the recordings and stimulations through two implanted electrodes, thus providing a constant connection between the two sites in the brain. The device had to be



small, but possess enough battery power to run for several days, and enough computational power to record signals and analyse the data in real time, according to Mavoori.

The chip also had to be sufficiently stable to function in living animals. "Monkeys do not sit very still. They don't seem to care if you are trying to do science," laughs Fetz, who worried that such movements would create too much background noise. But once implanted, the chip worked. "I remember the first time we downloaded the data from the chip and got a glimpse of a day in the life of a brain cell," recalls Jackson. "It was very exciting."

They next asked whether the microchip was having any effect on brain functions. At the start of the experiment, they used an external stimulator to test different sites in the monkeys' motor cortex. As expected, stimulation of site A caused wrist muscles to twitch in one direction, and stimulation of site B caused twitching in another. Then the monkeys were left to go about their business with the chip on their brain, where it recorded and transmitted signals. After a couple of days,

the researchers once again tested different areas of the brain with the external stimulator. They found that now stimulation of site A was causing the wrist to move in the same direction as stimulation of site B. Thus the artificial connection between A and B had changed A's function to be more like B's.

"I was rather surprised when we got the result," says Jackson. "We saw a change of output from a whole region of the cortex and this caused a lasting change in the movement produced by stimulation." This change in function, called plasticity, is thought to occur normally during learning and memory. The researchers speculate that in people who have suffered brain damage, the chip could be used to help reorganize information flow and boost the functioning of spared neurons. ■

KEY COLLABORATOR

In lieu of exact species counts, conservation strategies typically rely on the assumption that geographic diversity patterns are the same among all taxonomic groups. A 16-strong team of young conservation biologists, many already friends, decided to test that assumption.

Former housemates and PhD labmates Richard Grenyer and David Orme kept in touch while doing their respective postdocs. Grenyer was working on a massive mammal

database at the University of Virginia in Charlottesville, while Orme, based at Imperial College London, UK, focused his attention on a global distribution of birds. Inspired by the compatible data sets, they combined them with an amphibian database to form a huge inventory of more than 19,000 species. On page 93, the team reports that whereas the distribution of overall species richness is similar among the three groups,

there is little similarity in the distribution patterns of rare and threatened species.

"All of the data coming online at the same time was critical," says Grenyer, now a researcher at the Royal Botanic Gardens in Kew, London. But the key to this probably controversial work, he adds, was working with friends. "If you're working with people you don't know, or who are senior to you, your creativity may be constrained. You don't take risks," he says. ■